



University of Strathclyde
Glasgow

On the Simultaneous Inference of Susceptibility Distributions and Non-Pharmaceutical Interventions from Epidemic Trajectories

Ibrahim Mohammed

Department of Mathematics and Statistics
University of Strathclyde
Glasgow, UK

July 23, 2026

This thesis is submitted to the University of Strathclyde for award of the degree of Doctor
of Philosophy in the Faculty of Science.

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Parts of Chapters 2 and 3 are based on the published article [1] and the contribution to the Discussion reported in [2]. These chapters also contain substantial additional material developed specifically for this thesis, including further mathematical arguments, simulation studies, sensitivity analyses, and discussion that do not appear in those publications. The work was undertaken in collaboration with Chris Robertson and Gabriela M. Gomes. Parts of Chapters 4 and 5 draw on [3], an archived preprint that had been submitted for publication but was not yet published at the time of submission of this thesis. These chapters likewise contain additional methodological development, numerical analyses, and discussion beyond the material included in the preprint. This work was undertaken in collaboration with Chris Robertson and Gabriela M. Gomes.

The author contributed to all aspects of the research reported in [1,3]. For the contribution to the Discussion published as [2], the author undertook the coding and simulation studies and contributed to the analysis and preparation of the results, but did not draft or present the contribution.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Professor Gabriela M. Gomes and Professor Chris Robertson, for their invaluable guidance, patience, and constant encouragement throughout this research. Their expertise, insight, and support have been fundamental to the completion of this thesis.

I am especially grateful to my sponsor, the Petroleum Technology Development Fund (PTDF), Nigeria, for awarding me the scholarship that made this research possible. I also sincerely thank my employer, Abubakar Tafawa Balewa University, Bauchi (ATBU), for granting me permission and fellowship support to undertake this research. I am grateful to Professor Bala Maaji Abdulhamid and Professor A. B. Samaila of ATBU, Bauchi, for their encouragement and support throughout this journey. I also wish to acknowledge Alhaji Adamu Muhammad of PTDF, Nigeria, and Alhaji Sani Abba Udubo of the Central Bank of Nigeria for their assistance in facilitating the payment of my stipends during the course of my studies.

I am also grateful to the Department of Mathematics and Statistics at the University of Strathclyde for providing a supportive and stimulating environment in which to conduct my research. I extend my appreciation to my fellow postgraduate researchers and office colleagues for their friendship, encouragement, and assistance in many different ways. My thanks also go to the academic, administrative, and support staff who assisted me during my time at the University.

On a personal note, I owe an immeasurable debt of gratitude to my late father, Muhammad Lawal, for the sacrifices he made in support of my education and aspirations. To my mother, thank you for your endless love, prayers, and encouragement, which have continued to guide and sustain me.

Most importantly, I would like to thank my wife, Khadija, and my daughter, Fatimah, for their sacrifices, patience, perseverance, and unwavering support during my long absence. I am deeply grateful for their prayers, understanding, and constant belief in me.

Abstract

Individual variation in susceptibility can slow an epidemic through selective depletion of highly susceptible individuals, while non-pharmaceutical interventions (NPIs) reduce transmission through changes in contact behaviour. Because both mechanisms can produce similar epidemic slowdowns, distinguishing their effects from aggregate incidence data alone is difficult. This creates a practical identifiability problem for heterogeneous-susceptibility SEIR models and can lead to biased inference, misleading attribution of epidemic decline, and unreliable forecasts.

This thesis studies inference in SEIR models with heterogeneous susceptibility and time-varying NPIs, with emphasis on identifiability, forecasting, and robustness to modelling assumptions. Using simulation experiments, it first compares heterogeneous and homogeneous susceptibility models under both homogeneous and heterogeneous data-generating settings. The results show that the heterogeneous model can reproduce homogeneous-like behaviour when the population is effectively homogeneous, while providing more reliable inference and forecasts when susceptibility variation is present. In contrast, a homogeneous model may fit the observed part of an epidemic while misattributing selective depletion to stronger interventions and misforecasting subsequent dynamics.

The thesis then shows that single-epidemic inference is strongly affected by compensation between susceptibility heterogeneity and intervention strength. Across multiple inference methods and alternative NPI parameterisations, the coefficient of variation in susceptibility and the sustained contact level exhibit a pronounced trade-off, indicating that one epidemic trajectory often contains insufficient information to estimate both reliably. To address this, a joint-epidemic framework is developed in which related epidemics share common parameters but differ in their initial conditions. Joint fitting introduces additional contrasts in epidemic growth and susceptible depletion, thereby constraining the compensating directions present in single-epidemic analysis and substantially improving practical identifiability.

To examine robustness to the assumed form of susceptibility variation, the thesis develops discretisation methods for continuous-trait models beyond the standard Gamma formulation and

compares Gamma and Lognormal susceptibility families, with a homogeneous model included as a reference. The results show that distributional shape becomes important at moderate to high heterogeneity: although matched on mean and coefficient of variation, the Gamma and Lognormal models can produce different peak magnitudes, peak times, final sizes, parameter estimates, and forecasts. A misspecified heterogeneous model may still fit the observed epidemic well by shifting the estimated level of heterogeneity, and joint fitting can convert this hidden misspecification into precise but biased inference. Nevertheless, across the scenarios examined, using the wrong heterogeneous family is generally less damaging for forecasting than ignoring susceptibility heterogeneity altogether. The framework is finally applied to first-wave COVID-19 mortality data from England and Scotland using joint hierarchical models with different NPI specifications. The joint analysis produces informative estimates of susceptibility heterogeneity alongside country-specific intervention effects, but the inferred magnitude of heterogeneity remains sensitive to the assumed NPI form, the infection-to-death delay, the prior structure, and the fact that only two epidemics are available. Overall, the thesis shows that separating susceptibility heterogeneity from NPI effects remains difficult from a single epidemic, but that joint inference from related epidemic time series, combined with explicit assessment of model misspecification, provides a more informative and transparent framework for inference and forecasting in heterogeneous epidemic models.

Parts of the simulation studies reported in this thesis have been published in [1, 2]; related work underlying parts of the distributional misspecification analysis is archived as the preprint [3] and submitted for publication.

Contents

Acknowledgements	ii
Abstract	iii
1 Background, Literature Review, and Modelling Framework	1
1.1 Background and motivation	1
1.2 Classical Compartmental Epidemic Models	3
1.2.1 The SIR model	3
1.2.2 The SEIR model	4
1.2.3 Prevalence, force of infection, and incidence	5
1.3 Reproduction numbers and herd immunity	6
1.3.1 The basic reproduction number	7
1.3.2 The effective reproduction number	8
1.3.3 Herd immunity and the homogeneous threshold	9
1.4 Homogeneous and structured epidemic models	10
1.4.1 Age-structured models and WAIFW matrices	13
1.4.2 Risk-group and HIV/STI models	14
1.4.3 Network and agent-based models	15
1.4.4 Nonlinear incidence and power-law transmission	17
1.4.5 Richer compartmental epidemic models	18
1.5 Heterogeneous susceptibility models	20
1.5.1 Two kinds of susceptible individuals	20
1.5.2 Several kinds of susceptible individuals	21
1.5.3 Continuous susceptibility distributions	23
1.6 Gamma-distributed susceptibility models	25
1.6.1 Mean-one Gamma-distributed susceptibility	25

1.6.2	The heterogeneous susceptibility SEIR model	28
1.6.3	Reduction under Gamma-distributed susceptibility	30
1.6.4	Gamma as a shape-preserving susceptibility distribution	33
1.6.5	Reproduction number and heterogeneous herd-immunity threshold	36
1.7	Non-Pharmaceutical Interventions and Time-Varying Transmission	37
1.7.1	Autonomous and non-autonomous epidemic systems	39
1.7.2	Parametric NPI functional forms	39
1.7.3	Stringency-driven contact profiles	42
1.8	The Combined Heterogeneity–NPI Identifiability Problem	43
1.9	Research Gap and Thesis Contribution	46
1.10	Aim and Objectives	47
1.11	Thesis Outline	48
2	Heterogeneity in Susceptibility and NPIs: Parameter Identifiability	49
2.1	Introduction	49
2.2	Mathematical Model and NPIs	51
2.2.1	Transmission model	51
2.2.2	Baseline NPI parameterisation	51
2.2.3	Choice of epidemiological parameters	53
2.3	Structural and Practical Identifiability	54
2.3.1	Structural Identifiability: Input–Output Reasoning	56
2.3.2	Origin of the ν – c^* compensation ridge	59
2.4	Baseline Analysis: Practical Identifiability	60
2.4.1	Simulations	60
2.4.2	Typical epidemic trajectories under the four scenarios	61
2.5	Maximum likelihood estimation	62
2.6	Baseline analysis	63
2.6.1	Homogeneous datasets without NPIs (S1)	65
2.6.2	Homogeneous datasets with NPIs (S3)	65
2.6.3	Heterogeneous datasets without NPIs (S2)	66
2.6.4	Heterogeneous datasets with NPIs (S4)	67
2.6.5	Uncertainty quantification in S3 and S4 across replicated datasets	68
2.6.6	Visualising practical non-identifiability along the ν – c^* ridge	71
2.7	Robustness to stochastic process noise	72

2.7.1	Event-based stochastic simulation	74
2.7.2	Simulation design	75
2.7.3	Results	75
2.7.4	Precision gains with increasing population size	78
2.8	Epidemic trajectory forecasts	79
2.9	Chapter Summary and Conclusions	81
3	Joint-Epidemic Inference and Parameter Identifiability	83
3.1	Introduction	83
3.2	Mathematical Model and NPIs	85
3.3	Methodology: The Joint-Epidemic Framework	86
3.3.1	Joint Likelihood Function	86
3.3.2	Profile Likelihood Analysis	87
3.3.3	Eigenvalue and Hessian analysis	88
3.3.4	Simulation Design: The Large vs. Small Epidemic	88
3.4	Results: Profile likelihood and Hessian Analysis single vs. Two Epidemics	89
3.4.1	Quantitative Improvements: Precision and Bias	90
3.4.2	Numerical Stability and Condition Number	90
3.4.3	Parameter Correlations	91
3.4.4	Visualising the compensation ridge: bivariate likelihood surfaces and profile likelihoods	92
3.5	Testing Limits: The Three-Epidemic Extension	95
3.5.1	Simulation Design	96
3.5.2	Results: Diminishing Returns Two vs. Three Epidemics	97
3.6	Chapter Summary, Discussion and Conclusion	99
3.6.1	Summary of results	99
3.6.2	Discussions and Additional results	100
3.6.3	Conclusion	102
4	Distribution Discretisation Methods	103
4.1	Introduction and Motivation	103
4.2	Distribution-Agnostic Moment Algorithms	107
4.2.1	Split-multiplicative (SM) transformation	109

4.2.2	Log–Affine (LA) transformation: algebra, existence/uniqueness, implementation	112
4.3	Gamma Discretisation: Distribution-specific Ingredients	117
4.4	Numerical illustration for the Gamma discretisation	118
4.5	Trajectory Matching	123
4.6	Chapter Summary and Conclusion	127
4.6.1	Summary	127
4.6.2	Additional results from Appendix E	129
4.6.3	Conclusion	129
5	Distribution Misspecification in Heterogeneous Susceptibility Models	130
5.1	Introduction	130
5.2	Trajectory comparison and analysis	134
5.2.1	Transmission model used in this chapter	135
5.2.2	Discretisation and model initialization	137
5.2.3	Trajectory comparison metrics	137
5.2.4	Simulation grid	138
5.3	Trajectory comparison of Gamma vs Lognormal distributions using Fixed mean and CV	139
5.3.1	The pattern of divergence of trajectories	139
5.3.2	The impact of discretisation (k) and convergence	140
5.4	Distribution Misspecification Analysis	143
5.4.1	Data generation and observation model	143
5.4.2	Inference target and likelihood	144
5.4.3	Data-generating and fitted models	144
5.5	Inference for Lognormal-distributed susceptibility	145
5.5.1	Single-epidemic inference: Lognormal baseline	145
5.5.2	Fit and forecast for the Lognormal baseline (single epidemic)	147
5.5.3	Two-epidemic inference: Lognormal baseline	148
5.5.4	Fits and forecasts in joint inference (Lognormal baseline)	151
5.5.5	Asymmetry across the two baselines	152
5.6	Chapter Summary, Discussion and Conclusion	153
5.6.1	Discussion and Summary	153
5.6.2	Conclusion	155

6	Model-Based Assessment of Heterogeneity in the UK COVID-19 Epidemic	156
6.1	Introduction	156
6.2	Data alignment	157
6.3	Transmission model	157
6.3.1	Reduced SEIR with heterogeneous susceptibility	157
6.4	NPI parameterisations	159
6.4.1	Piecewise-linear contact reduction	159
6.4.2	Logistic contact reduction	160
6.4.3	OxCGRT stringency-driven contact profile	160
6.5	Observation model: from infections to deaths	161
6.5.1	Daily infections	161
6.5.2	Delay Convolution	161
6.6	Bayesian Inference Framework	161
6.6.1	Likelihood and posterior distribution	162
6.6.2	Parameter vector and hierarchical structure	163
6.6.3	Posterior predictive checks and fit metrics	165
6.7	Results	166
6.7.1	Model Fit To Observed Mortality	166
6.7.2	Posterior Summaries Of Shared Population-level Parameters	168
6.7.3	Country-specific posterior summaries	170
6.7.4	Model Comparison and Selection	174
6.7.5	Observation-layer sensitivity: infection-to-death delay	174
6.8	Comparison fits using uniform priors	179
6.8.1	Model specification for uniform priors	179
6.8.2	Posterior summaries of uniform prior fits	180
6.8.3	Comparison with the hierarchical model analysis	180
6.8.4	Comparison with Gomes et al. [4]	182
6.9	Discussion and Conclusions	183
6.9.1	Summary results: hierarchical models	183
6.9.2	Limitations and recommendations	184
6.9.3	Conclusions	186
7	General Summary and Conclusion	187
7.1	Overview	187

7.2	Summary of the main findings	188
7.2.1	Single-epidemic inference has clear practical limits	188
7.2.2	Joint-epidemic inference can substantially improve identifiability	188
7.2.3	Discretisation is necessary for moving beyond the reduced Gamma model	189
7.2.4	Distributional shape matters once heterogeneity is moderate to high	189
7.2.5	The empirical England–Scotland analysis supports partial but not complete separation	190
7.3	Main contributions of the thesis	191
7.4	Limitations of the present work	192
7.5	Directions for future work	193
7.6	Final conclusion	194
A	Mathematical and Statistical Details	195
A.1	Reference notation for the heterogeneous-susceptibility SEIR model	195
A.1.1	Continuous distribution model	195
A.1.2	Reduced Gamma reference form	196
A.1.3	Initial conditions: the E/I ratio in the early phase	197
A.1.4	Computing $\mathcal{R}_0$ for system (A.1)	197
A.2	Three NPI functional forms and window matching	199
A.2.1	The three functional forms	200
A.2.2	Matching the transition window	201
A.2.3	Numerical illustration	202
A.3	Basics of Bayesian Hierarchical Models	202
A.4	Sensitivity to the choice of epidemiological parameters	204
A.4.1	Sensitivity to $\mathcal{R}_0$	204
A.4.2	Joint sensitivity to $\mathcal{R}_0$ and ν	205
B	Robust ν-c^* Correlation Checks across Estimation Methods and NPI Profiles	206
B.1	Introduction	206
B.2	Simulation Design	207
B.3	Estimation Methods	207
B.4	Results	208
B.4.1	Summary of Correlations	209
B.4.2	Visual Evidence of the Compensation Ridge	209

B.4.3	Comparison with the analytical compensation argument	210
B.5	Conclusion	211
C	Supplementary Material for Chapters 2 and 3	212
C.1	Introduction	212
C.1.1	Profile Likelihood Results Across NPI Strengths	213
C.1.2	Bivariate likelihood surfaces across NPI strengths	215
C.1.3	Profile likelihood curves for the remaining shared parameters	215
C.2	Sensitivity to initial conditions and intervention strength	216
D	Higher-Dimensional Parameter Spaces via Bayesian Inference	221
D.1	Scenarios, Simulations, Model and Priors	221
D.2	Baseline Case: 4-Parameter Model	223
D.2.1	Posterior Summaries and precision Gains	223
D.2.2	Fit quality and Diagnostics: Baseline	223
D.2.3	Correlations and the ν - c^* ridge: Baseline	225
D.3	Scenario A:5-parameter Model	226
D.3.1	Posterior summaries: Scenario A	228
D.3.2	Fit quality and diagnostics: Scenario A	228
D.3.3	Correlations: scenario A	229
D.4	Scenario B:6-parameter Model	231
D.4.1	Posterior summaries: Scenario B	232
D.4.2	Fit quality and diagnostics: Scenario B	233
D.4.3	Re-emergence of the ν - c^* ridge: Scenario B	233
D.5	Appendix conclusion	235
E	Extension of Discretisation Methodology and Distribution-Level Validation	238
E.1	Introduction and Motivation	238
E.2	The Lognormal Distribution with unit mean	238
E.2.1	Representative Value Selection	240
E.3	Lognormal discretisation: distribution-specific ingredients	241
E.4	Numerical illustration for the Lognormal discretisation	242
E.5	Distribution-Level Validation	244
E.5.1	Overview	244
E.5.2	Density Error Metric	246

E.5.3	Numerical Stability at High Variance	246
E.5.4	Aggregate Summary of Density Errors	247
E.5.5	Error Reduction with Increasing Resolution	247
E.6	Summary	249
F	Supplementary results: distribution misspecification inference for the Gamma baseline	250
F.1	Inference for Gamma-distributed susceptibility	250
F.1.1	Single-epidemic inference: Gamma baseline	250
F.1.2	Fit and forecast for the Gamma baseline (single epidemic)	251
F.1.3	Two-epidemic inference: Gamma baseline	253
F.1.4	Fit and forecast for joint inference (Gamma baseline)	256
G	Supplementary Materials for Chapter 6	258
G.1	Observation-layer sensitivity and narrower-delay refit	258
G.1.1	Detailed posterior summaries for the narrower-delay logistic refit	259
G.2	MCMC Convergence Diagnostics	259
G.2.1	Hierarchical Model Diagnostics	260
G.3	Hierarchical Model Comparison and Selection	262
G.3.1	Three-Way Comparison	263
G.3.2	Hierarchical models pairwise Comparisons	264
G.3.3	Hierarchical Model Comparison Reliability	264
G.4	Fits with Uniform Priors	265
G.4.1	Model Fit and Residual Diagnostics	265
G.4.2	Posterior distribution and correlation plots comparisons	266

List of Figures

1.1	Flow diagram of the SIR model	4
1.2	Flow diagram of the SEIR model	5
1.3	Mean-one Gamma distribution for individual susceptibility	28
1.4	The herd immunity threshold	37
1.5	Heterogeneity vs Epidemic Trajectory	38
1.6	NPI functional forms plot	41
1.7	Illustration of local compensation between c^* and ν	46
2.1	Baseline NPI profile	52
2.2	Epidemic trajectories under four combinations of heterogeneity and NPIs	53
2.3	Effective transmission-reduction factor as a function of the representative suscep- tible fraction	60
2.4	Typical epidemic trajectories under heterogeneity and NPIs	62
2.5	Marginal MLE distributions for Scenario S3	66
2.6	Model fits for NPI scenarios S3 and S4	67
2.7	Marginal MLE distributions for Scenario S4	68
2.8	Joint distributions of replicated maximum-likelihood estimates	69
2.9	Fitted trajectories along the ν - c^* profile-likelihood ridge	72
2.10	Process-noise trajectories from the tau-leaping simulator	73
2.11	Representative fit to a stochastic tau-leaping trajectory	76
2.12	Joint MLE distributions from stochastic tau-leaping fits	77
2.13	Forecast performance under homogeneous model (Scenario S3)	80
2.14	Forecast performance under heterogeneous model (Scenario S4)	81
3.1	Two-epidemic incidence trajectories and linear NPI profile	89
3.2	Single versus two-epidemic joint distributions of profile-likelihood estimates . . .	92
3.3	Bivariate likelihood surfaces for the ν - c^* parameter space	93

3.4	Likelihood ridge map for the ν - c^* compensation direction	94
3.5	Profile likelihoods for ν and c^*	95
3.6	Three-epidemic incidence trajectories and linear NPI profile	96
3.7	Hessian condition number distributions for two-epidemic vs three-epidemic fits .	98
4.1	Log-affine root function for the discretised Gamma distribution	119
4.2	Discretised Gamma approximation versus the true PDF after LA transformation	119
4.3	Discretised Gamma approximation versus the true PDF after SM transformation	120
4.4	Calibrated Gamma representatives under equal probability spacing	122
4.5	Calibrated Gamma representatives under linear probability spacing	122
4.6	Trajectory error reduction relative to k equals 10 by CV and transformation . . .	126
4.7	Trajectories of discretised and reduced infection curves at k equals 20	127
4.8	Full SEIR trajectory comparison at moderate heterogeneity with log-affine transformation	128
5.1	Gamma and Lognormal density comparison with identical mean and CV	135
5.2	Higher-order moment differences between Gamma and Lognormal distributions .	136
5.3	Gamma versus Lognormal incidence trajectories at high discretisation resolution	140
5.4	Epidemic outcome differences between Lognormal and Gamma susceptibility distributions	141
5.5	Gamma versus Lognormal incidence trajectories at moderate discretisation resolution	142
5.6	MLE distributions for single-epidemic inference: Lognormal baseline	146
5.7	Single-epidemic forecast trajectories: Lognormal baseline	148
5.8	MLE distributions for two-epidemic inference: Lognormal baseline	150
5.9	Two-epidemic trajectory fits and forecasts: Lognormal baseline	152
6.1	Daily COVID-19 deaths for England and Scotland over the first-wave window . .	158
6.2	Observation-model pipeline linking latent infections to observed deaths	162
6.3	England fitted trajectories and NPI profiles for the three hierarchical specifications	167
6.4	Scotland fitted trajectories and NPI profiles for the three hierarchical specifications	168
6.5	Residuals over time for the three hierarchical NPI models	169
6.6	Posterior comparison of the minimum contact level across NPI specifications . .	170
6.7	Posterior comparison of susceptibility heterogeneity across NPI specifications . .	171
6.8	Posterior comparison of the basic reproduction number across NPI specifications	172
6.9	Observation-layer sensitivity from the fitted logistic incidence: England	176

6.10	Logistic fit with a narrower infection-to-death delay	177
6.11	Residual diagnostics for the logistic refit with a narrower infection-to-death delay	178
A.1	NPI functional forms: matched transition window	203
A.2	Sensitivity of daily incidence to the basic reproduction number $\mathcal{R}_0$	205
A.3	Joint sensitivity of daily incidence to $\mathcal{R}_0$ and ν , with $c^* = 0.3$	205
B.1	correlation vs methods of estimates vs NPIs	210
C.1	Single vs Two-epidemic models condition numbers across NPI levels	214
C.2	Off-baseline likelihood contours across NPI strengths	215
C.3	Profile likelihoods for $\mathcal{R}_0$ and t_0	216
C.4	Correlations vs NPI strength at $c^* = 0.4$	218
C.5	Correlations vs NPI strength at $c^* = 0.3$	219
C.6	Correlations vs NPI strength at $c^* = 0.2$	220
D.1	Posterior density comparison for the 4-parameter model	224
D.2	Residual diagnostics for the two-epidemic Bayesian fit	225
D.3	Fitted versus observed : two-epidemic Bayesian model	226
D.4	Trace plots for single vs two-epidemic 4-parameter model	227
D.5	Posterior densities for Scenario A	229
D.6	Residuals versus time Scenario A	230
D.7	Fitted versus observed daily incidence under the 5-parameter model (Scenario A)	231
D.8	Trace plots Scenario A	232
D.9	Posterior densities Scenario B	234
D.10	Fitted versus observed Scenario B	235
D.11	Residuals versus time :Scenario B	236
D.12	Trace plots :Scenario B	237
E.1	Log-Affine transformation	243
E.2	Fitted Log-normal Distribution by LA	244
E.3	Example 1: of 10 groups Discretised Log-normal Distribution	245
E.4	Example 2: of 10 groups Discretised Log-normal Distribution	245
E.5	Gamma Density Error Reduction	248
E.6	Lognormal Density Error Reduction for Log-normal	248
F.1	MLE distributions for single-epidemic inference: Gamma baseline	252

F.2	Single-epidemic forecast trajectories: Gamma baseline	253
F.3	MLE distributions for two-epidemic inference: Gamma baseline	255
F.4	Two-epidemic trajectory fits and forecasts: Gamma baseline	257
G.1	Trace plots for $\mathcal{R}_0$ across hierarchical NPI models	261
G.2	Trace plots for shared heterogeneity ν across hierarchical NPI models	262
G.3	Trace plots for the contact reduction parameter	263
G.4	Posterior correlation between ν and c^* in England	263
G.5	Posterior correlation between ν and c^* in Scotland	264
G.6	Pareto- k diagnostics for PSIS-LOO across hierarchical NPI models	266
G.7	Uniform-prior residuals over time	267
G.8	Uniform-prior posterior comparison of c^*	267
G.9	Uniform-prior fits and inferred $c(t)$ for England	268
G.10	Uniform-prior fits and inferred $c(t)$ for Scotland	268
G.11	Uniform-prior posterior comparison of ν	269
G.12	Uniform-prior posterior comparison of $\mathcal{R}_0$	269
G.13	Uniform-prior posterior correlation between ν and c^* in England	270
G.14	Uniform-prior posterior correlation between ν and c^* in Scotland	270

List of Tables

1.1	Selected literature on transmission structure, heterogeneity and epidemic control	11
1.2	Selected literature on transmission structure, heterogeneity and epidemic control (continued)	12
1.3	Evolution of susceptibility heterogeneity models	26
1.4	Key results linking heterogeneity, NPIs, and inferential uncertainty	44
2.1	Baseline simulation parameters motivated by COVID-19 dynamics	54
2.2	Baseline simulation–inference combinations.	64
2.3	Median inverse-Hessian parameter correlations for NPI scenarios	70
2.4	Uncertainty and bias summaries for baseline investigation	71
2.5	Parameter estimates from tau-leaping simulations	76
2.6	Reduction in confidence interval widths with increasing population size	78
3.1	Parameter estimates and confidence intervals from single vs two epidemics (NPI $c^* = 0.30$).	90
3.2	Hessian Matrix Condition Numbers Comparison	91
3.3	Inverse-Hessian parameter correlations for single- and two-epidemic fits	91
3.4	Parameter estimates and confidence intervals from two vs. three epidemics (NPI $c^* = 0.30$).	99
3.5	Inverse-Hessian parameter correlations for two- and three-epidemic fits	99
4.1	Log–affine transformation parameters and empirical checks ($CV \approx 1.155$, $k = 10$).	120
4.2	Split–multiplicative transformation parameters and empirical checks ($CV \approx 1.155$, $k = 10$).	121
4.3	Trajectory matching : Discretised vs reduced model with 20 groups of susceptibles.	124
4.4	Trajectory matching : Discretised vs reduced model with 10 groups of susceptibles.	125
4.5	Trajectory discrepancies by discretisation size	125

5.1	Peak incidence comparison: Gamma vs Lognormal at $k = 100$ and $k = 20$	143
5.2	Single-epidemic inference: Lognormal baseline	147
5.3	Single vs. two-epidemic inference under correct specification: Lognormal baseline	149
5.4	Two-epidemic inference: Lognormal baseline, all three fits	151
6.1	Goodness-of-fit metrics for hierarchical NPI models	167
6.2	Population-level posterior summaries across three NPI specifications	170
6.3	Country-level posterior summaries across three NPI specifications	173
6.4	Main-text PSIS-LOO and WAIC comparison across hierarchical NPI models . . .	174
6.5	Delay quantiles used in the logistic remapping sensitivity check	175
6.6	England delay-remapping sensitivity from fitted logistic incidence	176
6.7	Key parameter changes using the narrower-delay logistic refit	179
6.8	Shared heterogeneity using uniform priors	180
6.9	Country-level posterior summaries using uniform priors	180
6.10	Goodness-of-fit metrics for uniform priors	181
6.11	Key parameter comparison using linear NPI	182
A.1	NPI functional forms: window-matching formulas and simulation values	202
B.1	Robustness of ν - c^* correlation across estimation methods and NPI profiles	209
C.1	Correlation table across NPI strength levels	214
D.1	Bayesian posterior summaries: single vs two epidemics (4-parameter model) . . .	223
D.2	Selected posterior correlations: Bayesian 4-parameter model	226
D.3	Scenario A (5-parameter model posteriors)	228
D.4	Selected posterior correlations: Scenario A	230
D.5	Scenario B (6-parameter model posteriors)	233
D.6	Selected posterior correlations: Scenario B	235
E.1	LA transformation illustration	243
E.2	SM transformation illustration	243
E.3	LA and SM density error evaluation	247
F.1	Single-epidemic inference: Gamma baseline	251
F.2	Single vs. two-epidemic inference for the correct specification: Gamma baseline .	253
F.3	Two-epidemic inference: Gamma baseline, all three fits	254

G.1	Fixed inputs for Chapter 6 reduced mortality model	259
G.2	prior distribution for hierarchical models	260
G.3	Population-level posterior summaries for the logistic fit with shorter-spread delay	260
G.4	Country-level posterior summaries for the logistic fit with narrower delay	261
G.5	PSIS-LOO and WAIC comparison across hierarchical NPI models	264
G.6	Pairwise PSIS-LOO and WAIC comparisons	265
G.7	Prior distributions for centred models	265

Chapter 1

Background, Literature Review, and Modelling Framework

1.1 Background and motivation

Infectious diseases remain a major challenge for public health, food security, and social organisation. They affect human populations through diseases such as influenza, malaria, Ebola, HIV/AIDS, SARS, and COVID-19, but they also affect animal and plant populations. For this reason, epidemic modelling is not restricted to human medicine. It is also used in veterinary epidemiology, plant epidemiology, and broader One Health studies, where the health of humans, animals, plants, and the environment are treated as connected parts of the same system [5].

Respiratory infections provide one important motivation for this thesis. Seasonal influenza causes regular outbreaks, while pandemic respiratory viruses can spread rapidly and disrupt health services, education, work, travel, and economic activity. The COVID-19 pandemic gave a recent example of how a newly emerging respiratory virus can create large uncertainty for public-health planning. During such outbreaks, decision makers need to understand how quickly infection is spreading, what factors are shaping the observed epidemic pattern, and how reliable model-based projections are likely to be.

Mathematical models are useful because they make the assumptions behind epidemic reasoning explicit. A model does not reproduce every detail of a real population. Instead, it provides a simplified representation of the main processes thought to drive transmission. This simplification is useful because it allows different explanations of epidemic behaviour to be compared in a systematic way. Models can therefore be used to study possible epidemic scenarios, assess the consequences of assumptions, and interpret the information contained in observed data [6–8].

A central modelling tradition is the compartmental approach. In this approach, individuals are grouped according to their disease status, and the model describes how individuals move between these groups over time. This framework has its roots in early epidemic modelling, including the work of Hamer, Ross, and Kermack and McKendrick [9, 10].

Kleczkowski et al. [5] emphasise that this simple idea has developed into a broad modelling framework used across human, animal, and plant epidemics. They maintained that the strength of compartmental models is that they are simple enough to analyse but flexible enough to be extended. Depending on the scientific question, the same broad framework can be modified to include additional disease stages, age groups, risk groups, contact patterns, spatial structure, behavioural changes, or public-health measures. This flexibility is important because real epidemics are shaped by more than the biological properties of a pathogen. They are also shaped by the population in which the pathogen spreads and by the actions taken in response to the outbreak.

This thesis focuses on one part of this broader modelling problem: heterogeneity among individuals in a population and changes in transmission over time affect epidemic curves and the interpretation of epidemic data. The central problem is that observed epidemic decline can arise through different mechanisms. When susceptibility to infection is heterogeneous, infection is concentrated earlier among more susceptible individuals, so the average susceptibility of the remaining susceptible pool falls over time. In parallel, non-pharmaceutical interventions (NPIs), such as lockdown, reduce transmission through changes in contact behaviour. Since both mechanisms can flatten epidemic curves (see, e.g., [1, 4]), data from a single epidemic may not contain enough information to separate them reliably. This thesis studies that practical identifiability problem, develops computational tools for analysing it, and examines how far joint analysis across multiple epidemics can improve inference.

This chapter provides the background, literature review, and modelling framework for the thesis. Section 1.2 introduces classical compartmental epidemic models and defines the key quantities used throughout the thesis, including prevalence, force of infection, and incidence. Section 1.3 defines reproduction numbers and herd immunity, and explains why the classical homogeneous herd-immunity threshold is model-dependent. Section 1.4 reviews extensions of the compartmental framework that incorporate age structure, risk groups, network structure, nonlinear incidence, and richer compartmental formulations. Section 1.5 narrows the review to heterogeneous susceptibility models, moving from finite susceptibility groups to continuous susceptibility distributions. Section 1.6 introduces the Gamma-distributed susceptibility model

used in this thesis and derives the reduced heterogeneous SEIR formulation. Section 1.7 introduces time-varying transmission reductions through non-pharmaceutical interventions (NPIs). Section 1.8 brings the heterogeneity and NPI components together and identifies the central identifiability problem created when they are fitted simultaneously. Section 1.9 states the research gap and thesis contribution. Section 1.10 gives the aim and objectives. Finally, Section 1.11 outlines the structure of the remaining thesis.

1.2 Classical Compartmental Epidemic Models

Compartmental epidemic models describe disease transmission by dividing a population into epidemiological states and specifying how individuals move between those states over time. This framework dates back to the initial work of Hamer, Ross, and Kermack and McKendrick, and it remains central in mathematical epidemiology because it provides a simple but powerful way to describe epidemic growth, decline, and control [6–12].

Throughout this thesis, time is denoted by t . The main epidemiological compartments are

$$S(t), \quad E(t), \quad I(t), \quad R(t),$$

where $S(t)$ denotes the number of susceptible individuals, $E(t)$ the number of exposed individuals, $I(t)$ the number of infectious individuals, and $R(t)$ the number of removed individuals. The term *removed* means removed from the chain of transmission. Depending on the model and the data source, this may include individuals who have *recovered* and become immune, individuals who are effectively isolated, or individuals who have *died*.

For a closed population without births, non-disease deaths, or migration, the total population size is

$$N = S(t) + E(t) + I(t) + R(t),$$

and hence remains constant over time. This closed-population formulation is used in the main simulation chapters of the thesis because the epidemic windows considered there are short relative to demographic time scales.

1.2.1 The SIR model

The simplest standard compartmental model is the SIR model, which divides the population into *susceptible* (S), *infectious* (I), and *removed* (R) classes. In a closed population with homogeneous

mixing, the model is given by

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \quad (1.1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I, \quad (1.2)$$

$$\frac{dR}{dt} = \gamma I. \quad (1.3)$$

Here β is the transmission rate and γ is the removal rate, so that $1/\gamma$ is the mean infectious period. The term $\beta SI/N$ is described as frequency-dependent mass-action transmission. It represents the rate at which susceptible individuals become infected through contact with infectious individuals [7, 8].

Figure 1.1 shows the flow structure of the SIR model.

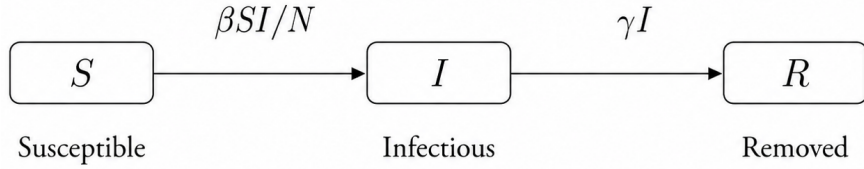


Figure 1.1: Individuals move from the susceptible class S to the infectious class I , and then to the removed class R .

The SIR model provides a useful foundation for epidemic theory, but it does not represent the interval between infection and entry into the infectious class. For many respiratory infections, including influenza and SARS-CoV-2, this delay is epidemiologically important and is represented by introducing an exposed compartment, leading to the SEIR model.

1.2.2 The SEIR model

The SEIR model extends the SIR framework by adding an exposed compartment $E(t)$ between infection and full infectiousness. The exposed class represents individuals who have been infected but have not yet entered the fully infectious class. In some applications, the exposed class is purely latent. In this thesis, following the model structure used by Gomes et al. [4], exposed individuals may contribute partially to transmission. Such a pre- I infectious stage is sometimes described as *prodromal* or *pre-symptomatic*, depending on the disease context.

The homogeneous SEIR model used as baseline notation in this thesis is

$$\frac{dS}{dt} = -\lambda(t)S(t), \quad (1.4)$$

$$\frac{dE}{dt} = \lambda(t)S(t) - \delta E(t), \quad (1.5)$$

$$\frac{dI}{dt} = \delta E(t) - \gamma I(t), \quad (1.6)$$

$$\frac{dR}{dt} = \gamma I(t). \quad (1.7)$$

Here $1/\delta$ is the mean exposed period before entry into the fully infectious class, and $1/\gamma$ is the mean infectious period. The force of infection $\lambda(t)$ is the per-susceptible instantaneous rate of infection. For the homogeneous model with partial infectiousness in the exposed class, we write

$$\lambda(t) = \beta \frac{\rho E(t) + I(t)}{N}, \quad (1.8)$$

where $\rho \in [0, 1]$ is the relative infectiousness of exposed individuals compared with fully infectious individuals. When $\rho = 0$, exposed individuals do not contribute to transmission. When $\rho = 1$, exposed and infectious individuals contribute equally. In the main simulation and data-analysis chapters, ρ is fixed rather than estimated.

Figure 1.2 shows the corresponding SEIR flow structure.

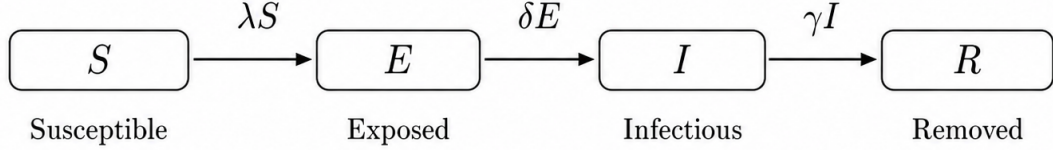


Figure 1.2: Flow diagram of the SEIR model. Individuals move from susceptible S to exposed E , then to infectious I , and finally to removed R . The transition rates are λS , δE , and γI , respectively.

1.2.3 Prevalence, force of infection, and incidence

It is important to distinguish prevalence, force of infection, and incidence, since these are related but different epidemiological quantities.

Prevalence is a stock variable: it counts how many individuals are in a given disease state at time t . For example, $I(t)$ is the prevalence of fully infectious individuals in the SEIR model. When exposed individuals also contribute to transmission, the effective infectious pool appearing in the force of infection is

$$\rho E(t) + I(t).$$

The force of infection $\lambda(t)$ is a hazard. It is the instantaneous rate at which a susceptible individual acquires infection at time t . Therefore, the total rate at which new infections occur is

$$\mathcal{I}_{\text{inf}}(t) = \lambda(t)S(t), \quad (1.9)$$

which is the flow from the susceptible class into the exposed class.

Incidence is the number of new cases per unit time [8]. In a compartmental model, the precise meaning of incidence depends on the event being counted. For the SEIR system, two related flow quantities are especially relevant. The first is the *infection incidence*,

$$\mathcal{I}_{\text{inf}}(t) = \lambda(t)S(t),$$

which counts newly infected individuals entering the exposed class.

The second is the *onset incidence*,

$$\mathcal{I}_{\text{onset}}(t) = \delta E(t), \tag{1.10}$$

which is the flow from the exposed compartment into the fully infectious compartment and therefore counts new entries into $I(t)$. This convention has also been used in SEIR analyses by introducing a cumulative case variable $C(t)$ satisfying $dC(t)/dt = \delta E(t)$, whose derivative is treated as the incidence curve, in for instance [13].

Throughout the simulation studies in Chapters 2–5, $\delta E(t)$ is used as the modelled incidence observable. It represents the onset of full infectiousness following the exposed period and provides a delayed model-based proxy for case onset. It should therefore be distinguished from the infection incidence, $\lambda(t)S(t)$, which records acquisition of infection and entry into the exposed compartment.

In Chapter 6, the observed data are daily deaths, so the latent incidence produced by the transmission model is taken to be the infection incidence, $\lambda(t)S(t)$, following Wood et al. [14], and is linked to deaths through an infection-to-death delay distribution and an observation model.

The next section introduces reproduction numbers and herd immunity. These concepts are built from the same compartmental framework, but they require separate treatment because herd immunity depends not only on the reproduction number, but also on the assumed structure of susceptibility in the population.

1.3 Reproduction numbers and herd immunity

Reproduction numbers and herd immunity are central threshold concepts in epidemic modelling. They provide a way to connect individual-level transmission assumptions to population-level epidemic growth, control, and vaccination thresholds. This section defines the basic reproduction

number, the effective reproduction number, and the classical homogeneous herd-immunity threshold. The heterogeneous herd-immunity threshold is introduced in section 1.6.5, after heterogeneity susceptibility SEIR models have been discussed.

1.3.1 The basic reproduction number

The basic reproduction number, denoted by $\mathcal{R}_0$, is the expected number of secondary infections generated by one typical infectious individual introduced into a wholly susceptible population [6–8]. It is a threshold quantity: if $\mathcal{R}_0 > 1$, an introduced infection can grow in a fully susceptible population, whereas if $\mathcal{R}_0 < 1$, sustained epidemic growth is not expected under the assumptions of the model [15, 16].

For the homogeneous SIR model in Section 1.2.1,

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \quad \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I,$$

the population is initially almost wholly susceptible, so $S(0) \approx N$. Near the disease-free state,

$$\frac{dI}{dt} \approx (\beta - \gamma)I.$$

Thus an epidemic initially grows if $\beta - \gamma > 0$, or equivalently if

$$\frac{\beta}{\gamma} > 1.$$

The basic reproduction number for this SIR model is therefore

$$\mathcal{R}_0 = \frac{\beta}{\gamma}. \tag{1.11}$$

This expression has a direct interpretation: β measures the rate at which an infectious individual generates infectious contacts in a fully susceptible population, and $1/\gamma$ is the mean infectious period.

For the homogeneous SEIR model used as the baseline notation in this thesis, the force of infection is

$$\lambda(t) = \beta \frac{\rho E(t) + I(t)}{N},$$

where $\rho \in [0, 1]$ is the relative infectiousness of exposed individuals compared with fully infectious individuals. An infected individual contributes to transmission during the exposed period, with

relative infectiousness ρ , and during the fully infectious period, with relative infectiousness one. The mean exposed period is $1/\delta$, and the mean infectious period is $1/\gamma$. Therefore, the basic reproduction number for this SEIR model is

$$\mathcal{R}_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right). \quad (1.12)$$

A formal derivation using the next-generation approach is given in appendix A.1.4, but the expression in (1.12) follows from adding the expected contribution to transmission during the exposed and infectious stages [4, 15, 16].

1.3.2 The effective reproduction number

The basic reproduction number refers to a wholly susceptible population. As an epidemic progresses, the number of susceptible individuals decreases, and so the transmission potential changes. The effective reproduction number, denoted here by $\mathcal{R}_{\text{eff}}(t)$, is the expected number of secondary infections generated by an infectious individual at time t , given the current state of the population [7, 8].

In a homogeneous model without interventions, the only population-level change affecting transmission is the remaining susceptible fraction,

$$s(t) = \frac{S(t)}{N}.$$

Therefore,

$$\mathcal{R}_{\text{eff}}(t) = \mathcal{R}_0 \frac{S(t)}{N}. \quad (1.13)$$

This formula says that transmission potential decreases as susceptible individuals are removed from the population through infection and recovery. If $\mathcal{R}_{\text{eff}}(t) > 1$, incidence tends to increase; if $\mathcal{R}_{\text{eff}}(t) < 1$, incidence tends to decrease, although the observed decline may occur with a delay because individuals already in the exposed or infectious compartments continue to progress through the system.

In section 1.7, non-pharmaceutical interventions are represented by a time-varying contact multiplier $c(t) \in (0, 1]$. In that case, the homogeneous effective reproduction number becomes

$$\mathcal{R}_{\text{eff}}(t) = \mathcal{R}_0 c(t) \frac{S(t)}{N}. \quad (1.14)$$

Equation (1.14) is not yet the full heterogeneous expression. It shows how susceptible depletion

and contact reduction enter the same threshold logic in the homogeneous setting.

1.3.3 Herd immunity and the homogeneous threshold

Definition. Herd immunity refers to population-level protection that arises when a sufficiently large fraction of individuals is immune, through previous infection or vaccination, so that the introduction of an infectious individual does not lead to sustained spread [8, 17]. Throughout this thesis the term is used in this *threshold* sense: herd immunity is reached when $\mathcal{R}_{\text{eff}}(t) < 1$ under the assumptions of the model. It does not mean that no further infections occur, nor that an epidemic stops immediately.

The phrase has been used in at least three distinct ways in the literature. Fine et al. [17] review this history and distinguish the proportion immune in a population, the threshold level of immunity needed for incidence to decline, and the indirect protection enjoyed by susceptible individuals when many others are immune. The threshold reading adopted here descends from classical epidemic theory: Kermack and McKendrick showed that epidemic growth depends on the balance between transmission and removal [10], and later work formalised that balance through the reproduction number [6, 8], with the next-generation framework supplying a general definition of $\mathcal{R}_0$ applicable to heterogeneous systems [15, 16]. In vaccination theory the same logic gives the critical coverage required to prevent sustained spread, under homogeneous mixing and perfect immunity [6–8].

The homogeneous threshold. In the homogeneous SIR or SEIR setting, suppose a fraction r of the population is immune, so that $s = 1 - r$ remains susceptible. Then $\mathcal{R}_{\text{eff}} = \mathcal{R}_0 s$, and the condition $\mathcal{R}_{\text{eff}} < 1$ gives $s < 1/\mathcal{R}_0$, hence $r > 1 - 1/\mathcal{R}_0$. For $\mathcal{R}_0 > 1$ the classical homogeneous herd-immunity threshold is therefore

$$H_{\text{hom}} = 1 - \frac{1}{\mathcal{R}_0}, \quad (1.15)$$

so that $\mathcal{R}_0 = 3$ requires roughly 66.7% immunity. Equation (1.15) makes the relationship to $\mathcal{R}_0$ explicit: the threshold is a deterministic, monotone function of the reproduction number alone, and is therefore a useful benchmark rather than a universal rule. It depends on assumptions including homogeneous mixing, equal susceptibility and exposure, permanent immunity, and a fixed transmission environment. Section 1.6.5 shows how relaxing equal susceptibility changes this threshold under Gamma-distributed susceptibility and compares the resulting heterogeneous threshold with the homogeneous benchmark, implying that it is model-dependent.

1.4 Homogeneous and structured epidemic models

The SIR and SEIR models introduced above are useful baseline models, but they rely on strong simplifying assumptions. In particular, the homogeneous model assumes that individuals are epidemiologically identical apart from their disease state, and that infectious contacts can be represented by a single average transmission rate. This is mathematically convenient, but real populations are structured. Individuals differ by age, contact behaviour, occupation, household structure, location, and biological response to infection. These differences can change both the speed of epidemic growth and the eventual size of an epidemic [5–7, 12]. There are several ways to relax the homogeneous-mixing assumption. Some models divide the population into a finite number of groups, such as age classes or risk groups, so that transmission depends on who mixes with whom rather than on one average contact rate [6, 7, 12, 18]. Other models represent contacts explicitly using networks, percolation ideas, small-world interaction rules, or spatial hierarchy [19–22]. A further line of work modifies the incidence term itself, for example by replacing the standard bilinear incidence term with a nonlinear or power-law expression [23–25]. These approaches are reviewed briefly here.

Tables 1.1 and 1.2 give a compact overview of the modelling themes that motivate this section. The table briefly show a progression from classical homogeneous compartmental models, through models on transmission scaling, age and risk structure, nonlinear incidence, networks, spatial hierarchy and intervention modelling, to COVID-era models with richer compartments, contact matrices, policy data, and high-dimensional intervention structure.

Three points are important. First, the transmission term is not merely algebraic bookkeeping: choices such as density-dependent transmission, frequency-dependent transmission, nonlinear incidence, or explicit mixing functions encode different biological and social assumptions. Second, homogeneous mixing can be relaxed in several mathematically distinct ways, including age and risk stratification, network topology, small-world contact structure, spatial hierarchy, and percolation-style descriptions of epidemic spread. Third, model structure matters for inference and control: assumptions about latent and infectious periods, contact reduction, and intervention timing can change estimates of transmission and the predicted impact of public-health measures.

We briefly give explicit model equations for selected types of structure discussed below, using them as simple representative examples rather than as full model specifications. The table supplies the broader literature map, while the equations show concretely how particular modelling assumptions enter the transmission dynamics to illustrate how different mechanisms can produce similar epidemic slowdowns, so an observed decline in incidence or deaths cannot

Table 1.1: Selected literature on transmission structure, heterogeneity and epidemic control

Paper	Theme	Main modelling contribution	Relevance to this thesis
Hamer [9]; Kermack and McKendrick [10]; Anderson and May [6]	Classical epidemic theory	Collectively establish the classical mass-action and compartmental epidemic framework.	Provides the homogeneous baseline against which the structured, heterogeneous, and time-varying transmission models in this thesis are compared.
Kleczkowski, Hoyle and McMenemy [5]	SIR/One Health review	Provides a cross-disciplinary foundation for viewing host heterogeneity and human intervention as distinct extensions of the classical SIR/SEIR framework, corresponding to the two mechanisms examined in this thesis.	Provides a cross-disciplinary foundation for viewing susceptibility heterogeneity and NPIs as distinct extensions of the classical SIR framework: the former reflects variation among hosts, while the latter represents changes arising from human action.
McCallum, Barlow and Hone [26]	Transmission scaling	Reviews density-dependent, frequency-dependent, and alternative transmission functions, showing that assumptions about how susceptible and infectious hosts interact can materially alter model behaviour.	Shows that the incidence function is a substantive modelling assumption, rather than a purely technical choice, and must be distinguished from the mechanistically derived nonlinear susceptible term used later in this thesis.
Busenberg and Castillo-Chavez [18]	Age/risk mixing and HIV	Develops an age- and risk-structured social/sexual mixing framework for HIV transmission, including consistency conditions for partner mixing and proportionate mixing as a special case.	Shows how heterogeneity in contact structure can be represented through explicit mixing functions rather than through a single homogeneous contact rate.
Liu, Hethcote and Levin [23]	Power-law incidence	Studies nonlinear incidence $\beta I^p S^q$ in SIS, SIR, SIRS, SEIS, SEIR, and SEIRS models, nonlinear incidence can produce multiple attraction basins.	Provides a classical example of phenomenological power-law incidence that is mathematically comparable to power-law susceptible term derived from Gamma-distributed susceptibility.
Hochberg [24]	Nonlinear transmission	Examines how nonlinear transmission rates alter the population dynamics of microparasitic host–pathogen systems.	Shows that the mathematical form of transmission can change epidemic dynamics and interpretation, while illustrating an alternative mechanism to the selective-depletion model used in this thesis.
Grassberger [19]	Spatial epidemic processes and percolation	Links a spatial stochastic general epidemic process to dynamical percolation and studies its critical behaviour near the epidemic threshold.	Provides historical background for spatial stochastic and percolation approaches that move beyond homogeneous mean-field ODE models.
Newman [20]	Networks, transmissibility and PGFs	Develops analytical solutions for classes of SIR epidemics on random networks using transmissibility, percolation, and probability-generating functions.	Provides an analytic network perspective that relaxes homogeneous mixing without relying only on agent-based simulation.
Kleczkowski and Grenfell [21]	Small-world mixing and effective contact rate	Studies a cellular-automaton epidemic model with local and global contacts, and shows that aggregated mean-field-type curves may fit well if the contact rate is treated as an effective parameter.	Shows that an aggregate epidemic curve can sometimes be approximated by a mean-field model with an effective contact rate, even when the underlying contacts are spatially structured.
Grenfell, Bjørnstad and Kappey [22]	Spatial hierarchy and epidemic waves	Uses measles data from England and Wales and a stochastic spatial model to show recurrent waves spreading from large core cities to smaller satellite towns; vaccination restricted but did not eliminate this hierarchy.	Demonstrates that population geography, spatial coupling, stochastic importation, and vaccination can shape the timing and spatial organisation of epidemic trajectories.
Keeling and Grenfell [27]	Infection-period distribution	Shows that the distribution of the infection period, defined as time spent in exposed and infectious classes, affects epidemic persistence and spatial spread.	Supports careful interpretation of latent and infectious period assumptions, including $1/\delta$, $1/\gamma$, and the broader role of model structure in inference.

Table 1.2: Selected literature on transmission structure, heterogeneity and epidemic control (continued)

Paper	Theme	Main modelling contribution	Relevance to this thesis
Wearing, Rohani and Keeling [28]	Model-structure bias	Shows that ignoring the latent period or assuming exponential latent and infectious periods can underestimate $\mathcal{R}_0$ from outbreak data and lead to over-optimistic control predictions.	Provides a close analogy for this thesis: misspecified model structure can bias transmission and intervention estimates.
Ferguson et al. [29]	Intervention modelling	Uses a spatially explicit influenza simulation model to evaluate targeted antiviral prophylaxis and social distancing for containing an emerging pandemic.	Provides a pre-COVID example showing that intervention timing, contact reduction, and implementation logistics matter for epidemic control.
Eubank et al. [30]	Realistic urban social networks	Models disease outbreaks on large-scale synthetic contact networks representing realistic urban social interactions.	Shows how explicit contact structure can generate epidemic dynamics that differ from homogeneous mixing assumptions.
Ferguson et al. [31]	Large-scale pandemic simulation	Uses large-scale individual-based simulation to evaluate mitigation strategies such as school closure, household quarantine, antiviral use, and social distancing.	Illustrates how detailed individual-level or spatial models can represent intervention logistics more explicitly than low-dimensional ODE models.
Dalgic et al. [32]	SEIR versus agent-based vaccine allocation	Compares vaccination strategies derived from deterministic SEIR modelling and agent-based simulation in a pandemic influenza setting.	Shows that different modelling resolutions can lead to different policy recommendations even under the same broad epidemic scenario.
Prem et al. [33]	Empirical contact matrices	Provides projected social contact matrices for many countries using contact surveys and demographic information.	Supplies the empirical basis for age-structured models in which transmission depends on who mixes with whom rather than on one average contact rate.
Youssef et al. [34]	Extended SEIR-type compartmental models	Expands the basic SEIR framework by adding extra disease or intervention compartments, allowing the model to represent additional clinical or public-health states.	Illustrates how the compartmental framework can be enlarged when the scientific question requires more detailed disease progression or intervention pathways.
Xu et al. [35]	Individual-based SEIHQR modelling	Uses an individual-based model with susceptible, exposed, infectious, hospitalised, quarantined and removed states to represent local COVID-19 transmission and contact-tracing interventions.	Shows that quarantine and hospitalisation can be represented explicitly, but also demonstrates that such detail increases model complexity and calibration demands.
Wang et al. [36]	Age-structured vaccination and reopening models	Uses an age-structured SEIR-type model with susceptible, exposed, asymptomatic, symptomatic, hospitalised, recovered and vaccinated compartments to assess vaccination and reopening strategies.	Shows how age-specific susceptibility, disease progression and vaccination can be incorporated when the main question concerns age-targeted control.
Britton et al. [37]	Age/activity heterogeneity and herd immunity	Uses age and activity structure to show that population heterogeneity can reduce the disease-induced herd-immunity threshold relative to the homogeneous benchmark.	Provides a structured-population comparison to the continuous susceptibility framework used in this thesis.
Ngonghala et al. [38]	High-dimensional COVID-era intervention models	Develops a compartmental model with vaccination, boosting, waning immunity, treatment, hospitalisation and multiple variants.	Demonstrates the flexibility of compartmental models for policy-rich questions, while motivating the more parsimonious model used in this thesis for identifiability analysis.
Hale et al. [39]	Policy data and stringency	Introduces OxCGRT, a panel database with policy indicators and composite indices, including the Stringency Index.	Supports the stringency-driven $c(t)$ model used later in Chapter 6.

automatically be attributed to either intrinsic selective depletion or extrinsic intervention.

1.4.1 Age-structured models and WAIFW matrices

Age is one of the most common forms of population structure in infectious disease modelling. For many directly transmitted infections, especially childhood infections such as measles, both susceptibility and contact patterns vary with age. Children, adults, and older individuals do not mix uniformly, and the risk of infection in one age group depends on infection prevalence in other age groups [6, 7, 12, 40].

A standard way to represent this structure is to divide the population into m age groups. Let

$$S_i(t), \quad E_i(t), \quad I_i(t), \quad R_i(t),$$

denote the numbers of susceptible, exposed, infectious, and removed individuals in age group i , for $i = 1, \dots, m$. The force of infection experienced by susceptible individuals in group i can be written as

$$\lambda_i(t) = \sum_{j=1}^m \beta_{ij} \frac{I_j(t)}{N_j}, \quad (1.16)$$

where N_j is the size of group j , and β_{ij} represents transmission from infectious individuals in group j to susceptible individuals in group i . The corresponding SEIR-type system is

$$\frac{dS_i}{dt} = -\lambda_i(t)S_i(t), \quad (1.17)$$

$$\frac{dE_i}{dt} = \lambda_i(t)S_i(t) - \delta_i E_i(t), \quad (1.18)$$

$$\frac{dI_i}{dt} = \delta_i E_i(t) - \gamma_i I_i(t), \quad (1.19)$$

$$\frac{dR_i}{dt} = \gamma_i I_i(t). \quad (1.20)$$

Here $1/\delta_i$ is the mean exposed period in group i , and $1/\gamma_i$ is the mean infectious period in group i . In many applications these rates are assumed to be common across groups, but the grouped formulation allows them to differ if the data or biology require it.

The matrix

$$\beta = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mm} \end{pmatrix}$$

is a transmission or mixing matrix. In epidemiology, such matrices are often called WAIFW

matrices, meaning “Who Acquires Infection From Whom” [6, 7, 12]. They generalise the single transmission parameter β in the homogeneous model. The entry β_{ij} describes transmission to group i from group j , so the model can represent assortative mixing, where individuals mainly mix within their own group, or disassortative mixing, where contacts occur mainly between different groups.

1.4.2 Risk-group and HIV/STI models

A second classical example of structured epidemic modelling comes from sexually transmitted infections (STIs), including HIV. For these infections, transmission depends strongly on partnership patterns and contact rates. A homogeneous model with one average transmission rate can hide the fact that a group with high contact rates may contribute disproportionately to epidemic spread [7, 40].

As a simple example, suppose the population is divided into a high-risk group H and a low-risk group L . Let S_H, I_H and S_L, I_L denote the susceptible and infectious proportions in the two groups. A simple risk-structured SIS-type model can be written as

$$\frac{dI_H}{dt} = \beta_{HH}S_HI_H + \beta_{HL}S_HI_L - \gamma I_H, \quad (1.21)$$

$$\frac{dI_L}{dt} = \beta_{LH}S_LI_H + \beta_{LL}S_LI_L - \gamma I_L. \quad (1.22)$$

Here β_{HH} describes transmission to the high-risk group from the high-risk group, β_{HL} transmission to the high-risk group from the low-risk group, and similarly for β_{LH} and β_{LL} . The four transmission terms are collected in the WAIFW matrix

$$\beta = \begin{pmatrix} \beta_{HH} & \beta_{HL} \\ \beta_{LH} & \beta_{LL} \end{pmatrix}. \quad (1.23)$$

This matrix plays the same conceptual role as the single parameter β in the homogeneous model, but it contains information about both contact rates and mixing between risk groups [7].

For m risk groups, the same idea gives

$$\frac{dI_i}{dt} = S_i \sum_{j=1}^m \beta_{ij} I_j - \gamma I_i, \quad i = 1, \dots, m. \quad (1.24)$$

If n_i is the proportion of the population in group i , then at the disease-free state $S_i = n_i$. The

next-generation matrix has entries of the form

$$K_{ij} = \frac{n_i \beta_{ij}}{\gamma_j}. \quad (1.25)$$

The basic reproduction number is then

$$\mathcal{R}_0 = \text{spr}(K), \quad (1.26)$$

where $\text{spr}(K)$ denotes the spectral radius, or dominant eigenvalue, of K .

This class of models is important for two reasons. First, it shows that heterogeneity in contact behaviour can change invasion thresholds and endemic prevalence. Second, it shows that the average behaviour of the population may be a poor guide to epidemic risk when transmission is concentrated in particular groups. This point is especially relevant for HIV/STI models, where risk structure, pair formation, and partnership duration may be central to the transmission process [7, 40].

1.4.3 Network and agent-based models

Age-structured and risk-group models still assume homogeneous mixing within each group. Network and agent-based models take a finer-scale approach by representing contacts, individuals, or local interaction settings more explicitly. The two terms are related, but they are not identical. A network model is primarily a model of contact structure: individuals or groups are represented as nodes and potentially infectious contacts as edges. An agent-based model is primarily an individual-level simulation framework: individuals are represented as agents with attributes, behaviours, locations, and transition rules. Many agent-based models are network-based, but an agent-based model may also use spatial movement, household-location mixing, or rule-based encounters rather than a fixed contact graph.

If A_{ij} is the adjacency matrix of a contact network, with $A_{ij} = 1$ when individuals i and j are connected and $A_{ij} = 0$ otherwise, then a simple network version of the force of infection on susceptible individual i is

$$\lambda_i(t) = \tau \sum_{j=1}^m A_{ij} Y_j(t), \quad (1.27)$$

where τ is the transmission rate per infectious edge, and $Y_j(t) = 1$ if node j is infectious at time t , with $Y_j(t) = 0$ otherwise. This formulation replaces the population-average infectious prevalence in the homogeneous model with the number of infectious neighbours of individual i .

Exact stochastic epidemic models on networks are usually high-dimensional. If each of N nodes can be in one of several disease states, the full state space grows exponentially with N . Therefore, exact models are generally too large for routine inference or analytic work. Several approximation frameworks have been developed to reduce this complexity [41, 42].

One family of approximations uses pairwise models. Instead of tracking only the number of susceptible and infectious individuals, pairwise models also track the number of connected pairs, such as $[SI](t)$, the expected number of susceptible–infectious edges. These models improve on homogeneous mixing because they retain information about who is connected to whom. They usually require closure approximations, because equations for pairs depend on triples, equations for triples depend on quadruples, and so on [41].

A second family uses degree-based or heterogeneous mean-field models. Let $[S_k](t)$ denote the expected number of susceptible nodes with degree k , where degree means the number of contacts of a node. Such models retain information about variation in the number of contacts between individuals. This is important because individuals with high degree are both more likely to become infected and more likely to transmit infection to others.

A third family uses probability-generating functions and edge-based compartmental models. In these approaches, the degree distribution of the network is summarised through a generating function, and epidemic spread is described in terms of the probability that transmission has not yet occurred along a randomly chosen edge. These models can give low-dimensional descriptions of SIR epidemics on large random networks and are closely related to percolation methods [20, 41].

Agent-based models are closely related, but they are usually simulation-based rather than analytically reduced. In these models, disease spread is generated through the simulated behaviour and interactions of individual agents. Agents may belong to households, workplaces, schools, or other locations; they may have age, risk status, vaccination status, quarantine status, or other attributes; and their contacts may change through time. This makes agent-based models useful when the timing, location, and repeated nature of contacts are important [30, 31, 35].

Network and agent-based models are not used as the main inferential model in this thesis, but they are relevant background. They show that heterogeneity in contacts, locations, and individual behaviour can strongly alter epidemic dynamics. They also provide one route by which individual-level variation can generate population-level epidemic effects that are not captured by a homogeneous model. The present thesis takes a more parsimonious route: it represents individual variation through a continuous susceptibility distribution, while keeping the epidemic model low-dimensional enough for practical identifiability analysis.

1.4.4 Nonlinear incidence and power-law transmission

A further way to improve homogeneous model and is to modify the incidence term itself. In the homogeneous SIR or SEIR model, the usual incidence term is bilinear in the susceptible and infectious populations, for example

$$\lambda(t) = \frac{\beta S(t)I(t)}{N}.$$

This form assumes that the rate of new infections is proportional to both the number of susceptible individuals and the number of infectious individuals, after the appropriate normalisation by population size. However, the bilinear form is not the only possible modelling choice. If contact saturation, spatial clustering, repeated exposure, behavioural avoidance, or heterogeneous mixing affects transmission, then the incidence rate may increase more slowly or more rapidly than the bilinear expression suggests.

A classical formulation of this idea was given by Liu, Hethcote and Levin [23].

They considered nonlinear incidence terms of the form

$$\lambda(t) = \beta I(t)^p S(t)^q, \quad (1.28)$$

where $\beta > 0$ is transmission rate, $p > 0$ and $q > 0$. The standard bilinear incidence term is recovered when $p = q = 1$. When either exponent differs from one, the incidence rate is no longer linear in the corresponding compartment. For example, $p < 1$ may represent saturation in the effect of infectious individuals, whereas $p > 1$ may represent situations in which repeated exposure or clustering makes incidence increase faster than linearly with infectious prevalence. Similarly, $q \neq 1$ changes how the remaining susceptible pool contributes to new infections.

Liu et al. [23] embedded this incidence term in a family of epidemic models, including SIS, SIR, SIRS, SEIS, SEIR and SEIRS systems. A representative SEIRS version can be written as

$$S' = -\beta I^p S^q + \mu - \mu S + \omega R, \quad (1.29)$$

$$E' = \beta I^p S^q - (\delta + \mu)E, \quad (1.30)$$

$$I' = \delta E - (\gamma + \mu)I, \quad (1.31)$$

$$R' = \gamma I - (\omega + \mu)R, \quad (1.32)$$

with

$$S + E + I + R = 1. \quad (1.33)$$

Here μ is the birth and death rate, δ is the exposed-to-infectious progression rate, γ is the recovery rate, and ω is the rate at which recovered individuals lose immunity.

The main lesson from Liu et al. [23] is that nonlinear incidence can change the qualitative behaviour of epidemic models. Bilinear incidence models often have a relatively simple threshold structure, with either a disease-free equilibrium or a stable endemic equilibrium depending on parameter values. By contrast, models with incidence $\beta I^p S^q$ can have multiple equilibria, multiple basins of attraction, and periodic solutions generated through Hopf bifurcations. Thus, whether infection dies out or persists may depend not only on parameter values, but also on initial conditions. This is important because it shows that the mathematical form of the incidence term can change the interpretation of epidemic decline, persistence, and control.

Hochberg [24] made a related point from the perspective of transmission efficiency. Instead of assuming a constant transmission coefficient, the transmission term can be written in a more general form in which transmission efficiency depends on host densities. In simplified notation, the bilinear term βIS is replaced by

$$\beta(S, I) IS,$$

so that the per-contact or per-density efficiency of transmission is itself allowed to vary with the densities of susceptible and infected hosts. This reinforces the same modelling lesson: the incidence term is not merely algebraic bookkeeping, but a biological assumption about how contacts, susceptibility, infectiousness and population structure combine to generate new infections.

These classical nonlinear-incidence models are directly relevant to the present thesis because the reduced heterogeneous-susceptibility model introduced in section 1.6.3 also produces a power-law modification of the susceptible term.

1.4.5 Richer compartmental epidemic models

The SEIR model is a standard starting point, but the compartmental idea is much broader than the four classes S, E, I, R . Depending on the scientific question, researchers often refine the state space to represent quarantine, hospitalisation, vaccination, age structure, testing, multiple variants, or explicit contact heterogeneity. A few examples are useful here because they show the breadth of compartmental modelling before narrowing to the framework used in this thesis, which focuses on one particular direction of generalisation: heterogeneity in susceptibility.

Quarantine, isolation and hospitalisation compartments: the SEIHQR model

One common extension is to introduce explicit quarantine and hospitalisation compartments. This is useful when the model is intended to represent case detection, clinical progression, separation of infectious individuals from the general population, and recovery under supervision. Xu et al. [35] developed an individual-based SEIHQR model for local COVID-19 transmission, in which each person moves stochastically among susceptible, exposed, infectious, hospitalised, quarantined, and removed states.

In this model, the quarantine compartment Q is driven by contact tracing rather than by ordinary disease progression. Once a case is confirmed and hospitalised, close contacts, and in stricter scenarios second-layer contacts, are traced and placed in quarantine. Infected quarantined contacts may later be diagnosed and transferred to hospital, whereas uninfected quarantined contacts return after the quarantine period. Thus Q represents supervised isolation of traced contacts, while H represents confirmed cases removed from the contact network through hospitalisation. The key addition relative to a basic SEIQR is the explicit hospitalisation state H : confirmed cases are removed from the contact network by hospitalisation, and the quarantine compartment captures contacts placed under supervised isolation before confirmation. This separation allows the model to distinguish between clinical outcomes and public health containment actions, and to study how the reach and speed of contact tracing alter epidemic dynamics. The broader point is that once policy questions about detection, hospitalisation, and tiered quarantine are central, the basic SEIR state space becomes too small to represent the intervention structure faithfully.

High-dimensional COVID-era intervention models

COVID-19 also generated larger compartmental models that combine vaccination, boosting, waning immunity, variant structure, testing, hospitalisation, and treatment. For example, Ngonghala et al. [38] developed a two-variant Delta–Omicron model with shared susceptible classes and variant-specific exposed, infectious, hospitalised, and recovered states. Such models are useful when the scientific question concerns combined intervention packages, healthcare burden, or variant competition. However, they are substantially more parameter-rich than the SEIR models used in this thesis, and therefore introduce additional inferential complexity.

These examples show that the compartmental framework is broad and adaptable. Models can be expanded by adding new disease states, by stratifying the population into interacting groups, or by moving to individual-based simulation. However, each added layer of structure

introduces new parameters and new assumptions. This thesis does not attempt to model all forms of structure at once. Instead, it focuses on one specific mechanism: continuous heterogeneity in susceptibility, combined with time-varying NPIs. This choice keeps the model mechanistic and sufficiently tractable for identifiability analysis, and it aligns directly with the main research question of the thesis, namely whether heterogeneity in susceptibility and contact reduction can be distinguished reliably from epidemic time-series data. For that reason, the rest of this chapter moves toward the line of work most directly relevant here: models in which susceptibility is treated as a population distribution, especially the Gamma frailty framework, for example [4, 43, 44].

1.5 Heterogeneous susceptibility models

Section 1.4 shows that the homogeneous model can be generalised in several ways. This section narrows the focus to the single form of heterogeneity used throughout the thesis: variation in susceptibility to infection. The premise is that susceptible individuals need not experience the same infection hazard under identical epidemic conditions, whether because of biological factors, prior immune history, vaccination status, or other unobserved risk factors. Differences in exposure or contact behaviour can generate heterogeneous risk as well, but unless stated otherwise the individual trait modelled here is susceptibility.

The literature on heterogeneous susceptibility can be read as a progression that relaxes the homogeneous assumption in stages: from dividing the susceptibles into two groups, to several discrete classes, and finally to a continuous trait described by a probability distribution. The subsections that follow trace this route, from finite susceptibility groups to the Gamma-distributed model used in the remainder of the thesis.

1.5.1 Two kinds of susceptible individuals

One of the earliest treatments of susceptibility variation is due to Gart [45], who introduced a deterministic model with two susceptible sub-populations and later extended it to several groups within a chain-binomial framework [46]. These early models were applied successfully to epidemic data on yaws and chickenpox, showing that even a simple split of the susceptible population can be of practical value.

In the two-group version, let $S_1(t)$ and $S_2(t)$ denote the numbers of susceptible individuals in each group, and let $I(t)$ denote the total number of infectious individuals, comprising the initial infectives together with those generated from both groups. In Gart’s simple-epidemic setting, infectious individuals are not removed during the observed period, and the susceptible equations

are

$$\frac{dS_1}{dt} = -\beta_1 I(t) S_1(t), \quad (1.34)$$

$$\frac{dS_2}{dt} = -\beta_2 I(t) S_2(t), \quad (1.35)$$

where $\beta_1, \beta_2 > 0$ are the group-specific infection rates.

The infection pressure is generated by the total infectious population $I(t)$, so the two groups differ only through the rates β_1 and β_2 , not through who infects whom. This is the same structure as the risk-group models of Section 1.4.2: all infectives contribute to a common transmission pressure, while the susceptibles in each group are depleted at different rates. If $\beta_1 > \beta_2$, the first group is depleted faster, and the composition of the remaining susceptible population shifts towards the less susceptible group as the epidemic proceeds.

This is the discrete analogue of selective depletion. The higher-susceptibility group supplies a disproportionate share of the early infections, leaving a susceptible pool that is less susceptible on average. Using this framework, Gart [45] obtained exact and approximate solutions and estimated infection rates from epidemic data with two kinds of susceptibles.

1.5.2 Several kinds of susceptible individuals

The two-group idea extends naturally to several susceptibility classes. In Ball's [47] deterministic model the population is divided into J susceptibility groups. Let $S_j(t)$, $I_j(t)$, and $R_j(t)$ denote the susceptible, infectious, and removed populations in group j , for $j = 1, \dots, J$, and let

$$I_{\mathcal{D}}(t) = \sum_{j=1}^J I_j(t)$$

denote the total infectious population. The finite-group model is

$$\frac{dS_j}{dt} = -\alpha_j S_j(t) I_{\mathcal{D}}(t), \quad (1.36)$$

$$\frac{dI_j}{dt} = \alpha_j S_j(t) I_{\mathcal{D}}(t) - \gamma I_j(t), \quad (1.37)$$

$$\frac{dR_j}{dt} = \gamma I_j(t), \quad (1.38)$$

for $j = 1, \dots, J$, where $\alpha_j > 0$ is the infection-rate parameter for group j and γ is the common removal rate. As in the two-group case, transmission is driven by the total infectious population $I_{\mathcal{D}}(t)$, since infected individuals are assumed equally infectious; the groups differ in susceptibility alone.

This is a finite-dimensional system, and the total susceptible population is the finite sum

$$S(t) = \sum_{j=1}^J S_j(t), \quad (1.39)$$

rather than an integral over a continuous trait; the integral formulation appears only once the finite set of classes is replaced by a continuous susceptibility distribution (Section 1.5.3). Ball [47] showed that replacing such a heterogeneous population by a uniform one with the same average susceptibility can overstate the predicted spread of infection. This matches the intuition from the two-group model: because the most susceptible individuals are infected earliest, the average susceptibility of those who remain falls, so a model built on the initial average susceptibility overestimates onward transmission.

Hyman and Li [48] placed this idea on a firmer analytical footing, formulating differential-susceptibility SIR models in which the susceptible population is split into multiple subgroups by susceptibility. They derived explicit reproduction numbers, established the stability of the disease-free and endemic equilibria, and showed that differential susceptibility can be treated within a systematic compartmental framework. The approach was later enriched with further control and demographic mechanisms: Zhang et al. [49] added time delay and pulse vaccination, and Zhang et al. [50] added stage structure and pulse vaccination. Such extensions accommodate vaccination, delay, and stage structure, but at the cost of a larger and more parameter-rich model.

A related but distinct finite-class model was proposed by Nakata and Omori [51]. Although it again uses only two susceptible classes, it is grouped here rather than with the static two-group model of Section 1.5.1 because its defining feature is that susceptibility is not fixed: individuals change class after infection through repeated infection and the waning or boosting of immunity. With $0 < \sigma_2 < \sigma_1 \leq 1$, their model is

$$\frac{dS_1}{dt} = -\beta\sigma_1 S_1 I + \gamma\theta I, \quad (1.40)$$

$$\frac{dS_2}{dt} = -\beta\sigma_2 S_2 I + \gamma(1 - \theta)I, \quad (1.41)$$

$$\frac{dI}{dt} = \beta\sigma_1 S_1 I + \beta\sigma_2 S_2 I - \gamma I, \quad (1.42)$$

where $\theta \in [0, 1]$ is the probability of entering the higher-susceptibility class after recovery. This is not the framework used in the thesis, since the simulation chapters assume removal after

infection, but it is useful background because it treats susceptibility as a dynamic individual property rather than a fixed one. Nakata and Omori showed that reallocating susceptibility through repeated infection can produce delayed outbreaks, in which the infectious population first declines and only later grows.

Finite-group models make susceptibility variation explicit while retaining a compartmental structure, and in this chapter they serve mainly as a bridge from homogeneous mixing to the continuous distribution used in the thesis model. Their drawback is that the modeller must fix the number of groups, the susceptibility of each group, and the initial group proportions in advance. When susceptibility is not naturally divided into a few known classes, a continuous distribution provides a more parsimonious representation, and it is to that formulation we now turn.

1.5.3 Continuous susceptibility distributions

The continuous formulation was developed explicitly by Novozhilov [52], who considered a closed heterogeneous population in which each individual carries a fixed trait value that affects transmission. For the susceptibility case this trait is the individual's susceptibility to infection. Although the trait itself is fixed, its distribution among those still susceptible changes over time, because individuals with larger susceptibility are infected earlier.

Let $x \in \mathcal{X}$ denote the susceptibility trait and $s(t, x)$ the density of susceptible individuals at time t with trait value x , so that $s(t, x) dx$ is approximately the number of susceptibles with trait in $[x, x + dx]$. The total susceptible population is

$$S(t) = \int_{\mathcal{X}} s(t, x) dx. \quad (1.43)$$

This is a continuous-trait description rather than the finite-group formulations of Sections 1.5.1 and 1.5.2, with the sum over classes replaced by an integral over the trait.

Following Novozhilov [52], take the infectious population to be homogeneous with size $I(t)$, and let $\beta(x)$ be the transmission parameter associated with trait x . The distributed-susceptibility SIR model is

$$\frac{\partial s(t, x)}{\partial t} = -\beta(x)s(t, x)I(t), \quad (1.44)$$

$$\frac{dI(t)}{dt} = I(t) \int_{\mathcal{X}} \beta(x)s(t, x) dx - \gamma I(t), \quad (1.45)$$

$$\frac{dR(t)}{dt} = \gamma I(t), \quad (1.46)$$

with initial condition

$$s(0, x) = S_0 p_s(0, x), \quad (1.47)$$

where $p_s(0, x)$ is the initial probability density of susceptibility among susceptibles.

Writing the current susceptibility density among susceptibles as

$$p_s(t, x) = \frac{s(t, x)}{S(t)}, \quad (1.48)$$

and the mean transmission parameter among those still susceptible as

$$\bar{\beta}(t) = \int_{\mathcal{X}} \beta(x) p_s(t, x) dx, \quad (1.49)$$

the infectious equation becomes

$$\frac{dI(t)}{dt} = \bar{\beta}(t) S(t) I(t) - \gamma I(t). \quad (1.50)$$

The heterogeneous model therefore has the outward form of a homogeneous SIR model, except that the effective transmission coefficient $\bar{\beta}(t)$ is no longer constant: it declines over time as infection sculpts the susceptible distribution towards its less susceptible members.

Novozhilov [52] further showed that a distributed-susceptibility model can be reduced to a homogeneous model with a nonlinear transmission function whose form is fixed by the initial susceptibility distribution. When that distribution is Gamma, the nonlinear term is a power law, which supplies a mechanistic derivation of the widely used phenomenological incidence term

$$\beta S^a I^b.$$

In the susceptibility-only case the exponent on the susceptible term exceeds one and is set by the variance of the susceptibility distribution. This result is central to the thesis, because the reduced Gamma-susceptibility SEIR model used in later chapters inherits exactly this power-law structure.

The same framework yields the final size of an epidemic for an arbitrary initial susceptibility distribution [52], and Katriel [53] sharpened this result by characterising how the distribution controls the attack rate. He derived the attack-rate equation for an arbitrary susceptibility distribution and bounded the outcome on both sides: a sharp upper bound depending on $\mathcal{R}_0$ alone, and a sharp lower bound depending on $\mathcal{R}_0$ together with the variance. At fixed $\mathcal{R}_0$, a

larger variance of susceptibility distribution gives a smaller attack rate. The long-run behaviour of a heterogeneous model therefore depends on the whole initial distribution, not merely on its first one or two moments. The Gamma distribution is thus a convenient special case, chosen for analytic tractability, and the choice of heterogeneity distribution remains a modelling assumption in its own right.

The discussion has thus moved from finite susceptibility classes to continuous, frailty-style susceptibility distributions. Table 1.3 summarises this progression, showing how each step relaxes a different part of the homogeneous assumption and identifying the route that leads to the Gamma-distributed model used in this thesis. The next section specialises the continuous framework to the Gamma distribution, derives the reduced model used throughout, and explains how the power-law exponent in the macroscopic equations is related to coefficient of variation in equation (1.53).

1.6 Gamma-distributed susceptibility models

The previous section introduced heterogeneous susceptibility as a modelling idea. This section specialises that idea to the Gamma distribution and derives the reduced model used throughout the thesis. The purpose is to move from the susceptibility-indexed model, in which each compartment is indexed by an individual trait x , to a finite-dimensional SEIR system in which the effect of susceptibility heterogeneity is represented by the coefficient of variation ν .

The model follows the line of work developed by Novozhilov [52], Gomes et al. [4], Montalbán et al. [44], Rose et al. [43], and Neipel et al. [25]. The key assumption is that susceptibility is a positive individual trait. Individuals with larger susceptibility are infected earlier on average, so the distribution of susceptibility among the remaining susceptible individuals changes during the epidemic. When the initial susceptibility distribution is Gamma, this selective depletion has a particularly simple mathematical form.

1.6.1 Mean-one Gamma-distributed susceptibility

Let $X > 0$ denote individual susceptibility to infection. The interpretation is relative: an individual with $X = 2$ has twice the infection hazard of an individual with $X = 1$, under the same epidemic conditions. The susceptibility distribution is normalised to have mean one,

$$\mathbb{E}[X] = 1. \tag{1.51}$$

Table 1.3: Evolution of susceptibility heterogeneity models relevant to this thesis

Approach	Representative papers	Core idea or mechanism	Relevance to this thesis
Two susceptible classes	Gart [45]	The susceptible population is divided into two groups with different infection rates, while all infectives contribute to a common infection pressure.	Provides an early and transparent example of selective depletion: the higher-susceptibility group is infected earlier, so the composition of the remaining susceptible population changes over time.
Several susceptible classes	Gart [46]; Ball [47]; Hyman and Li [48]	Susceptible individuals are partitioned into finitely many groups with different susceptibility levels or infection hazards.	Extends the two-class idea to a finite number of risk groups. This captures heterogeneity more flexibly, but the number of groups and their susceptibility values must be specified in advance.
Differential susceptibility with additional epidemic mechanisms	Zhang et al. [49, 50]; Bonzi et al. [54]; Nakata and Omori [51]	Finite susceptibility classes are combined with additional mechanisms such as delay, pulse vaccination, stage structure, repeated infection, or changing susceptibility after recovery.	Shows that susceptibility heterogeneity can be embedded in richer compartmental systems, but also illustrates the increasing parameter burden created by finite-class extensions.
Continuous susceptibility distributions	Novozhilov [52]; Veliov [55]; Coutinho et al. [56]	Susceptibility is treated as a continuous positive trait. Individuals with larger susceptibility are infected earlier, so the mean susceptibility of the remaining susceptible pool decreases over time.	Provides the continuous-trait framework used in this thesis. It replaces a finite set of susceptibility groups with a probability distribution, but requires choosing a distributional family.
Frailty interpretation	Aalen et al. [57]; Coutinho et al. [56]; Novozhilov [52]	Individual susceptibility is interpreted as an unobserved multiplicative factor on the infection hazard, analogous to frailty in survival analysis.	Justifies treating susceptibility as a latent positive random variable and motivates the use of the coefficient of variation ν as a dimensionless summary of individual variation.
Gamma-distributed susceptibility	Novozhilov [52]; Rose et al. [43]; Gomes et al. [4]; Montalbán et al. [44]	The Gamma distribution is positive, analytically tractable, and shape-preserving under epidemic selection. With mean-one Gamma susceptibility, selective depletion leads to a reduced model with a power-law susceptible term.	This is the main heterogeneity model used in the thesis. The coefficient of variation ν determines the exponent $1 + \nu^2$ in the reduced SEIR infection term.
Heterogeneity and herd-immunity thresholds	Britton et al. [37]; Gomes et al. [4]; Montalbán et al. [44]	Variation in susceptibility, exposure, or contact activity can lower the disease-induced herd-immunity threshold relative to the homogeneous benchmark.	Connects the mathematical model to the epidemiological interpretation of epidemic slowdown. It also motivates why estimates of ν matter for interpreting final size and herd-immunity calculations.
Sculpting and kinetic-order interpretation	Rose et al. [43]; Neipel et al. [25]	The infection process sculpts the susceptibility distribution by preferentially removing high-susceptibility individuals. Gamma-like forms lead to power-law epidemic dynamics and self-limiting waves.	Provides the direct conceptual bridge between individual-level heterogeneity and population-level epidemic curves. It also explains why heterogeneous dynamics can mimic external mitigation in time-series fits.
Distributional alternatives and misspecification	Rose et al. [43]; Neipel et al. [25]; Montalbán et al. [44]	Other initial susceptibility distributions may behave differently from the Gamma benchmark, although some may be sculpted towards Gamma-like forms under epidemic selection.	Motivates the later robustness chapter: if the fitted susceptibility distribution differs from the true distribution, estimates of ν , c^* , and forecasts may change.

This convention fixes the scale of susceptibility. Without it, both the mean of X and the transmission parameter β would scale the force of infection, making their separate interpretation ambiguous. With the mean fixed at one, β keeps its baseline interpretation for an average-susceptibility population, while the spread of X represents heterogeneity.

Definition : The coefficient of variation(CV), ν , of the susceptibility distribution, X , is defined by the relation:

$$\nu = \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}[X]}. \quad (1.52)$$

Under the mean-one convention, this becomes

$$\nu = \sqrt{\text{Var}(X)}. \quad (1.53)$$

Throughout this thesis, the notation

$$X \sim \text{Gamma}(\alpha, \alpha)$$

uses the shape-rate parameterisation, with both the shape and rate parameters equal to $\alpha > 0$.

Thus, the mean-one Gamma susceptibility distribution is written as

$$X \sim \text{Gamma}(\alpha, \alpha). \quad (1.54)$$

Its density is

$$q(x) = \frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x}, \quad x > 0, \quad \alpha > 0. \quad (1.55)$$

Under this shape-rate parameterisation,

$$\mathbb{E}[X] = \frac{\alpha}{\alpha} = 1, \quad \text{Var}(X) = \frac{\alpha}{\alpha^2} = \frac{1}{\alpha}, \quad \nu^2 = \frac{1}{\alpha}. \quad (1.56)$$

Equivalently,

$$\alpha = \frac{1}{\nu^2}. \quad (1.57)$$

The mean-one convention is illustrated in Figure 1.3. The vertical dashed line marks the population-average susceptibility, fixed at $x = 1$. Values below this line correspond to individuals with below-average susceptibility, while the right tail represents individuals with above-average susceptibility. The right-skewed shape is epidemiologically important because individuals in the upper tail experience a larger infection hazard, $x\lambda(t)$, and are therefore infected

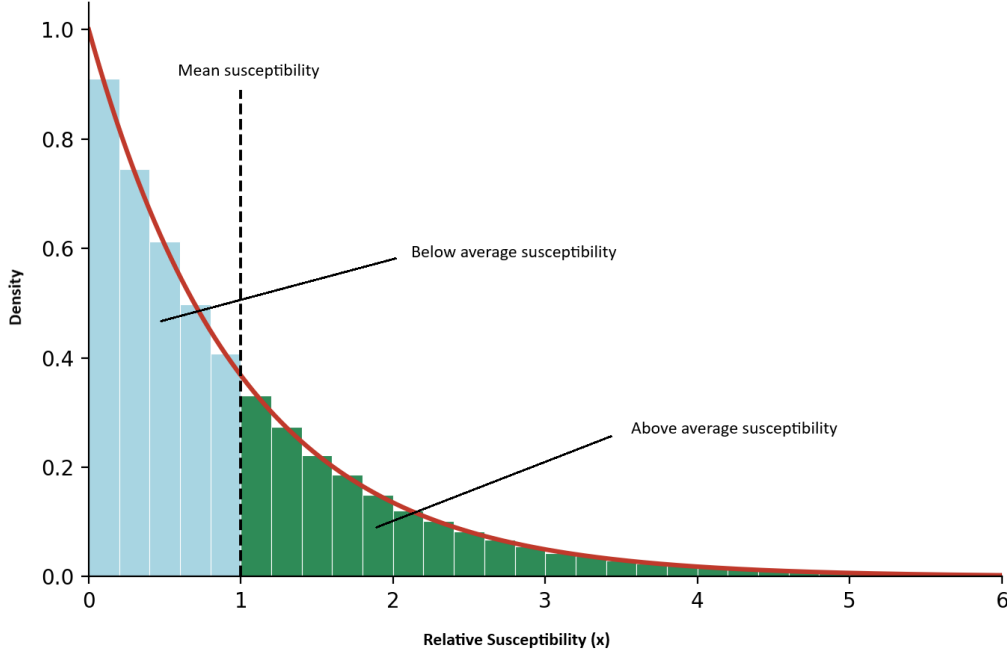


Figure 1.3: Illustration of a mean-one Gamma distribution for individual susceptibility. The dashed vertical line marks the population-average susceptibility, fixed at $x = 1$. Individuals with $x < 1$ have below-average susceptibility, while individuals in the right tail have above-average susceptibility. Because the infection hazard is proportional to $x\lambda(t)$, individuals in the right tail are infected preferentially early in the epidemic. This selective depletion progressively lowers the average susceptibility of the remaining susceptible pool.

preferentially early in the epidemic. Their selective removal changes the composition of the remaining susceptible pool and is the mechanism that produces the reduced Gamma-susceptibility model used below [43, 44, 52].

The homogeneous model is recovered in the limit $\nu \rightarrow 0$, corresponding to $\alpha \rightarrow \infty$, where the susceptibility distribution collapses to a point mass at $x = 1$.

The Gamma distribution is used because it is supported on positive values, can represent right-skewed individual variation, is standard in frailty modelling, and leads to an exact finite-dimensional reduction of the susceptibility-indexed epidemic model. It is viewed as a tractable benchmark, not as a claim that every real population has exactly Gamma-distributed susceptibility.

1.6.2 The heterogeneous susceptibility SEIR model

Let $S(x, t)$, $E(x, t)$, $I(x, t)$, and $R(x, t)$ denote the densities of susceptible, exposed, infectious, and removed individuals at time t with susceptibility x . The total numbers in each compartment are defined by

$$S(t) = \int_0^\infty S(x, t) dx, \quad E(t) = \int_0^\infty E(x, t) dx, \quad (1.58)$$

and

$$I(t) = \int_0^\infty I(x, t) dx, \quad R(t) = \int_0^\infty R(x, t) dx. \quad (1.59)$$

This notation is important.

Here $S(x, t)$ is a density with respect to the susceptibility trait x : $S(x, t) dx$ represents the number of susceptible individuals whose susceptibility lies in the interval $[x, x + dx]$. Consequently, the aggregate number of susceptible individuals at time t is

$$S(t) = \int_0^\infty S(x, t) dx. \quad (1.60)$$

The function $q(x)$ is the probability density of susceptibility in the initial susceptible population. Thus, if the initial susceptible population has size S_0 ,

$$S(x, 0) = S_0 q(x), \quad \int_0^\infty q(x) dx = 1. \quad (1.61)$$

Under the mean-one normalisation,

$$\int_0^\infty x q(x) dx = 1, \quad (1.62)$$

so that x measures susceptibility relative to the initial population average.

The quantity $S(x, t)$ is a population density with respect to the susceptibility trait x . Thus, $S(x, t) dx$ represents the number of susceptible individuals whose susceptibility lies in the interval $[x, x + dx]$. The aggregate susceptible population is therefore obtained by integrating $S(x, t)$ over all susceptibility values.

The function $q(x)$ specifies the probability density of susceptibility in the initial population and satisfies

$$\int_0^\infty q(x) dx = 1, \quad \int_0^\infty x q(x) dx = 1.$$

The second condition normalises the initial mean susceptibility to one, so that x represents susceptibility relative to the initial population average. If the initial susceptible population has size $S(0)$, then

$$S(x, 0) = S(0)q(x) \approx Nq(x), \quad (1.63)$$

where the approximation applies when the initial exposed and infectious populations are small relative to N . As the epidemic progresses, $S(x, t)$ evolves away from its initial form because individuals with larger values of x are infected more rapidly.

The susceptibility-indexed SEIR model is

$$\frac{\partial S(x, t)}{\partial t} = -x\lambda(t)S(x, t), \quad (1.64)$$

$$\frac{\partial E(x, t)}{\partial t} = x\lambda(t)S(x, t) - \delta E(x, t), \quad (1.65)$$

$$\frac{\partial I(x, t)}{\partial t} = \delta E(x, t) - \gamma I(x, t), \quad (1.66)$$

$$\frac{\partial R(x, t)}{\partial t} = \gamma I(x, t). \quad (1.67)$$

Here $1/\delta$ is the mean exposed period and $1/\gamma$ is the mean fully infectious period. As in the SEIR model introduced earlier, exposed individuals may contribute partially to transmission.

The force of infection is

$$\lambda(t) = \frac{\beta}{N} \int_0^\infty \{\rho E(x, t) + I(x, t)\} dx = \frac{\beta}{N} \{\rho E(t) + I(t)\}, \quad (1.68)$$

where $\rho \in [0, 1]$ is the relative infectiousness of the exposed class.

The infection hazard experienced by a susceptible individual of type x is $x\lambda(t)$. Thus all susceptibility classes experience the same epidemic environment through $\lambda(t)$, but the individual hazard is multiplied by x .

The basic reproduction number for the system (1.64)–(1.67) is

$$\mathcal{R}_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right). \quad (1.69)$$

Derivation of this expression is given in Appendix A.1.4.

As $q(x)$ follows a distribution with mean 1, the coefficient of variation ν , by definition in equation (1.53), is therefore given by

$$\nu = \sqrt{\int (x - 1)^2 q(x) dx}. \quad (1.70)$$

1.6.3 Reduction under Gamma-distributed susceptibility

The reduction follows from the cumulative force of infection. Define

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (1.71)$$

Solving (1.64) gives

$$S(x, t) = S(x, 0)e^{-x\Lambda(t)}. \quad (1.72)$$

Using the mean-one initial distribution $S(x, 0) \approx Nq(x)$, the susceptible fraction is

$$\frac{S(t)}{N} = \int_0^\infty q(x) e^{-x\Lambda(t)} dx. \quad (1.73)$$

This is the Laplace transform of the initial susceptibility distribution evaluated at $\Lambda(t)$, as in Novozhilov's distributed-susceptibility formulation [52].

For the gamma density in (1.55),

$$\begin{aligned} \frac{S(t)}{N} &= \int_0^\infty \frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x} e^{-x\Lambda(t)} dx \\ &= \frac{\alpha^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\alpha+\Lambda(t))x} dx \\ &= \left(\frac{\alpha}{\alpha + \Lambda(t)} \right)^\alpha. \end{aligned} \quad (1.74)$$

Using $\alpha = 1/\nu^2$, this becomes

$$\frac{S(t)}{N} = \left(1 + \nu^2 \Lambda(t) \right)^{-1/\nu^2}. \quad (1.75)$$

Let

$$s(t) = \frac{S(t)}{N}.$$

Then

$$s(t) = \left(\frac{\alpha}{\alpha + \Lambda(t)} \right)^\alpha.$$

Differentiating with respect to Λ gives

$$\frac{ds}{d\Lambda} = - \left(\frac{\alpha}{\alpha + \Lambda(t)} \right)^{\alpha+1} = -s^{1+1/\alpha} = -s^{1+\nu^2}. \quad (1.76)$$

Since $S(t) = Ns(t)$ and $d\Lambda/dt = \lambda(t)$,

$$\begin{aligned} \frac{dS}{dt} &= N \frac{ds}{d\Lambda} \frac{d\Lambda}{dt} \\ &= -N s^{1+\nu^2} \lambda(t) \\ &= -\beta \{ \rho E(t) + I(t) \} \left(\frac{S(t)}{N} \right)^{1+\nu^2}. \end{aligned} \quad (1.77)$$

Therefore, when susceptibility is Gamma-distributed with mean one and coefficient of variation ν , the infinite system (1.64)–(1.67) reduces to

$$\frac{dS}{dt} = -\beta(\rho E + I) \left(\frac{S}{N} \right)^{1+\nu^2}, \quad (1.78)$$

$$\frac{dE}{dt} = \beta(\rho E + I) \left(\frac{S}{N} \right)^{1+\nu^2} - \delta E, \quad (1.79)$$

$$\frac{dI}{dt} = \delta E - \gamma I, \quad (1.80)$$

$$\frac{dR}{dt} = \gamma I. \quad (1.81)$$

When $\nu = 0$, the power $1 + \nu^2$ becomes one and the standard homogeneous SEIR infection term is recovered:

$$-\beta(\rho E + I) \frac{S}{N}.$$

Thus the reduced model preserves the homogeneous model as a limiting case.

The power-law term in (1.78) is the mathematical measure of susceptibility heterogeneity in this thesis. It follows from the initial Gamma distribution and the selective depletion of high-susceptibility individuals. It elegantly captures the entire effect of Gamma-distributed heterogeneity in a single exponent, directly linking the microscopic population variance (ν^2) to the macroscopic epidemic dynamics to the broader nonlinear-incidence literature, where terms of the form $S^a I^b$ have been used to summarise departures from homogeneous mass-action transmission [25, 52].

Remark. For Gamma-distributed susceptibility with coefficient of variation ν , the reduced SEIR model contains the infection term

$$\beta(\rho E(t) + I(t)) \left(\frac{S(t)}{N} \right)^{1+\nu^2}. \quad (1.82)$$

This resembles the power-law incidence term in equation (1.28), introduced in Section 1.4.4, but its interpretation is different. In the phenomenological nonlinear-incidence model, equations (1.29)–(1.32), the exponents p and q are introduced directly to represent departures from bilinear transmission. In the Gamma-susceptibility model used in this thesis, the exponent $1 + \nu^2$ is derived from an underlying distribution of individual susceptibility. The exponent is therefore tied to a measurable population-level quantity, the coefficient of variation in susceptibility, rather than being only an empirical shape parameter.

The distinction is important for the thesis argument. Classical power-law incidence models show that modifying the infection term can change epidemic dynamics. The heterogeneous-susceptibility model gives one mechanistic route by which such a nonlinear term can arise: selective infection of more susceptible individuals lowers the average susceptibility of those remaining in the susceptible pool. This provides the bridge to the next subsection, which shows

explicitly why the Gamma distribution is shape-preserving under epidemic selection.

1.6.4 Gamma as a shape-preserving susceptibility distribution

Section 1.6.3 obtained the power-law infection term by direct calculation. This subsection explains *why* that calculation works: the Gamma distribution is shape-preserving under epidemic selection. The argument follows the susceptibility-sculpting framework of Rose et al. [43], the power-law heterogeneity formulation of Neipel et al. [25], and the Gamma-reduction framework of Montalbán et al. [44]. Everything below follows from a single object, the Laplace transform of the initial susceptibility density,

$$M(\Lambda) = \int_0^\infty q(x) e^{-x\Lambda} dx, \quad (1.83)$$

already encountered as the right-hand side of (1.73), where $\Lambda(t)$ is the cumulative force of infection (1.71). In the notation of Neipel et al. [25], Λ is the progress variable τ defined by $\dot{\tau} = \beta I/N$, and $s(t) = M(\Lambda(t))$ is their expression for the susceptible population; Rose et al. [43] use a dimensionless progress variable $\phi(t)$ for the same purpose.

Sculpting as an exponential tilt. From the solution (1.72) of the susceptibility-indexed susceptible equation, $S(x, t) = S(x, 0) e^{-x\Lambda(t)}$, the density of susceptibility among those who *remain* susceptible at time t is

$$q_t(x) = \frac{S(x, t)}{S(t)} = \frac{q(x) e^{-x\Lambda(t)}}{M(\Lambda(t))} \propto q(x) e^{-x\Lambda(t)}. \quad (1.84)$$

The epidemic therefore acts on the initial distribution by an exponential tilt. The factor $e^{-x\Lambda(t)}$ suppresses large x , so high-susceptibility individuals leave the susceptible pool earlier. This is the mechanism Rose et al. [43] identify as sculpting: the infection process and the susceptibility distribution shape one another, so that both the number of susceptible individuals and the average susceptibility of those still susceptible decline together.

Selective depletion holds for any q . Writing $\bar{x}(t) = \int_0^\infty x q_t(x) dx$ for the mean susceptibility of the remaining pool, (1.84) gives $\bar{x} = -M'(\Lambda)/M(\Lambda) = -\frac{d}{d\Lambda} \log M(\Lambda)$, whose derivative is minus the variance of the tilted law. Since $d\Lambda/dt = \lambda(t)$,

$$\frac{d\bar{x}(t)}{dt} = -\lambda(t) \text{Var}_t(X). \quad (1.85)$$

Thus, whenever $\lambda(t) > 0$ and susceptibility varies at all, the mean susceptibility of the remaining pool decreases, at a rate equal to the current variance. This is the general statement of selective depletion, and it requires no distributional assumption.

The Gamma case: shape preservation. Now take the mean-one Gamma density (1.55), $q(x) \propto x^{\alpha-1}e^{-\alpha x}$, with $\alpha = 1/\nu^2$ by (1.56). Substituting into (1.84) and discarding factors independent of x ,

$$q_t(x) \propto x^{\alpha-1}e^{-\alpha x}e^{-x\Lambda(t)} = x^{\alpha-1}e^{-\{\alpha+\Lambda(t)\}x},$$

which is an unnormalised Gamma kernel. Hence

$$X \mid \text{still susceptible at time } t \sim \text{Gamma}(\alpha, \alpha + \Lambda(t)). \quad (1.86)$$

The shape parameter is unchanged; only the rate increases, from α to $\alpha + \Lambda(t)$. Consequently

$$\bar{x}(t) = \frac{\alpha}{\alpha + \Lambda(t)}, \quad \text{Var}_t(X) = \frac{\alpha}{\{\alpha + \Lambda(t)\}^2}, \quad \nu_t = \frac{1}{\sqrt{\alpha}} = \nu. \quad (1.87)$$

The gamma law remains Gamma, with the same shape and the same coefficient of variation, but with a decreasing mean. Rose et al. [43] call this an *eigen-distribution* of the sculpting process; Neipel et al. [25] describe the same property as shape invariance and emphasise that it holds at all times, not merely as an initial condition. The mean in (1.87) is their Eq. (14) written with Λ in place of τ .

Why shape preservation gives the power law. Because the shape is preserved, the whole state of the susceptible pool is carried by the single scalar $\bar{x}(t)$. Comparing (1.87) with the susceptible fraction (1.75), $s(t) = \{\alpha/(\alpha + \Lambda)\}^\alpha$, gives $s(t) = \bar{x}(t)^\alpha$ and therefore

$$\bar{x}(t) = s(t)^{1/\alpha} = s(t)^{\nu^2}. \quad (1.88)$$

The susceptibility-weighted susceptible pool, which is what the force of infection actually acts on, is then

$$\int_0^\infty xS(x, t) dx = S(t)\bar{x}(t) = N \left(\frac{S(t)}{N} \right)^{1+\nu^2}, \quad (1.89)$$

and substituting into $-dS/dt = \lambda(t) \int_0^\infty xS(x, t) dx$ returns the reduced infection term (1.78). In the SIR notation of Rose et al. [43], the infection rate $\beta I \bar{x} S$ becomes $\beta I S^{1+1/\alpha}$, so the *kinetic*

order of the epidemic is

$$p = 1 + \frac{1}{\alpha} = 1 + \nu^2. \quad (1.90)$$

The kinetic-order reading of Rose et al. and the Gamma-reduction of Montalbán et al. [44] are therefore two descriptions of one mechanism. The homogeneous model is $p = 1$; an exponential susceptibility distribution, being Gamma with $\alpha = 1$, gives $p = 2$ [43].

Gamma as an attractor. Shape preservation is not confined to a Gamma initial condition. Neipel et al. [25] parameterise heterogeneity by the power-law exponent of the susceptibility density at *small* susceptibilities, $q(x) \sim Cx^{\alpha-1}$ as $x \downarrow 0$, the same α as the Gamma shape used here, and show that the Gamma law is the attractor of the dynamics for that whole class. The reason is visible in (1.84): for large Λ the tilt $e^{-x\Lambda}$ suppresses large x , so q_t is increasingly governed by the behaviour of q near the origin, and

$$q_t(x) \approx \frac{\Lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\Lambda x}, \quad (1.91)$$

a Gamma density with shape α and rate Λ . Every initial distribution sharing the same small- x exponent therefore shares the same limiting dynamics [25, 43]. This is an asymptotic statement, not a claim that every real epidemic runs long enough for the limit to be reached; Chapter 5 examines what happens when the assumed distribution is wrong.

Two consequences for inference. The results above are not only structural; they bear directly on the identifiability problem posed in Section 1.8. First, Rose et al. [43] show that models of different kinetic order are effectively indistinguishable during the early, exponential-like phase of an outbreak, precisely the phase in which epidemic models are most often parameterised, so that first-order fits may systematically overstate the eventual severity of an outbreak. Second, and more directly, Neipel et al. [25] report that changes in mitigation strength can be compensated to some degree by changes in the heterogeneity exponent, and conclude that quantifying the effect of mitigation requires knowing the degree of heterogeneity in the population. The compensation between ν and c^* studied in this thesis is therefore an instance of a difficulty already visible in the heterogeneity literature; what the later chapters add is a systematic identifiability analysis of it and an assessment of whether joint inference across epidemics can break it.

The sculpting framework used here, in common with Rose et al. [43], assumes no correlation between an individual's susceptibility and their propensity to transmit, so that highly susceptible individuals are not also assumed to be disproportionate spreaders. Relaxing that assumption

changes epidemic speed and final size [58] and is not pursued in this thesis.

1.6.5 Reproduction number and heterogeneous herd-immunity threshold

The basic reproduction number for the system (1.64)–(1.67) is given by (1.69). Heterogeneity does not change this initial value from that of equation (1.12); its effect appears later through selective depletion of highly susceptible individuals. In the absence of time-varying interventions,

$$\mathcal{R}_{\text{eff}}(t) = \mathcal{R}_0 \left(\frac{S(t)}{N} \right)^{1+\nu^2}. \quad (1.92)$$

Setting $\mathcal{R}_{\text{eff}} = 1$ gives the heterogeneous herd-immunity threshold

$$H_{\text{het}} = 1 - \mathcal{R}_0^{-1/(1+\nu^2)} = 1 - \mathcal{R}_0^{-\alpha/(\alpha+1)}, \quad \alpha = \frac{1}{\nu^2}. \quad (1.93)$$

The homogeneous threshold is recovered when $\nu = 0$:

$$H_{\text{hom}} = 1 - \frac{1}{\mathcal{R}_0}. \quad (1.94)$$

Thus $H_{\text{het}} < H_{\text{hom}}$ whenever $\nu > 0$. For $\mathcal{R}_0 = 3$, the threshold decreases from 66.7% under homogeneous susceptibility to 42.3% at $\nu = 1$ and 19.7% at $\nu = 2$ (Figure 1.4). When susceptibility varies between individuals, infection is not distributed uniformly: more susceptible individuals are infected earlier, so the average susceptibility of the remaining susceptible pool falls over time. This selective depletion brakes transmission more strongly than the homogeneous model predicts, and consequently (1.15) can *overstate* the disease-induced threshold [4, 37, 43, 44].

Figure 1.5 illustrates the differences between the dynamics of homogeneous (system (1.7)) and heterogeneous susceptibility SEIR model in the system (1.64)–(1.67). As ν increases, highly susceptible individuals are removed from the susceptible pool early, slowing epidemic growth intrinsically without any external intervention. For example, with $\mathcal{R}_0 = 3.0$ and $\nu = \sqrt{2} \approx 1.414$, roughly 50% of the population is infected over the course of the outbreak, compared with approximately 90% under the homogeneous assumption ($\nu = 0$). The bottom panel shows the corresponding incidence trajectories, with the homogeneous model producing the highest and sharpest peak when heterogeneity is unaccounted for.

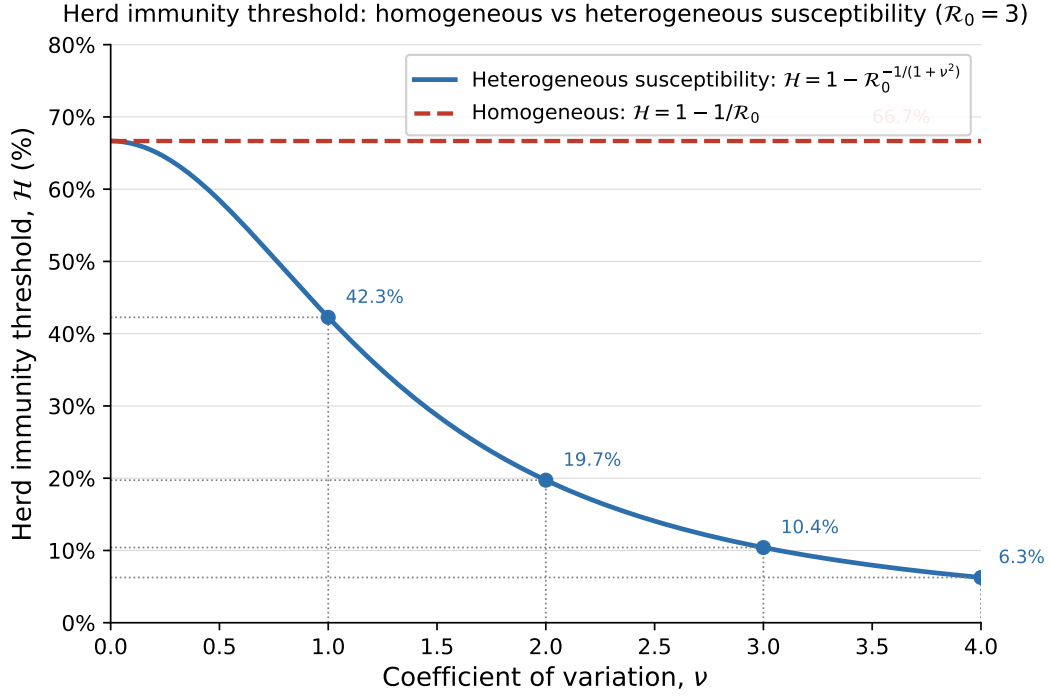


Figure 1.4: Herd immunity threshold $\mathcal{H}$ as a function of the coefficient of variation ν for $\mathcal{R}_0 = 3$. The dashed red line is the homogeneous threshold, $\mathcal{H}_{\text{hom}} = 1 - 1/\mathcal{R}_0 = 66.7\%$. The solid blue curve is the heterogeneous threshold, $\mathcal{H}_{\text{het}} = 1 - \mathcal{R}_0^{-1/(1+\nu^2)}$, which decreases as heterogeneity increases. Grey dotted lines mark the values at $\nu = 1, 2, 3, 4$.

1.7 Non-Pharmaceutical Interventions and Time-Varying Transmission

The previous section described epidemic slowdown that arises endogenously from susceptible depletion and, in the heterogeneous model, from selective depletion of individuals with higher susceptibility. Epidemic trajectories can also be shaped by exogenous changes in contact behaviour. During an outbreak, these changes may arise from voluntary behavioural responses, public-health advice, or mandated non-pharmaceutical interventions (NPIs), such as physical distancing, school and workplace closures, restrictions on gatherings, isolation of cases, quarantine of contacts, and stay-at-home orders.

In compartmental epidemic models, a common way to represent such changes is to allow the transmission rate to vary over time. In this thesis, NPIs are introduced through a contact multiplier $c(t)$, where

$$0 < c(t) \leq 1, \quad (1.95)$$

and

$$\beta(t) = \beta c(t). \quad (1.96)$$

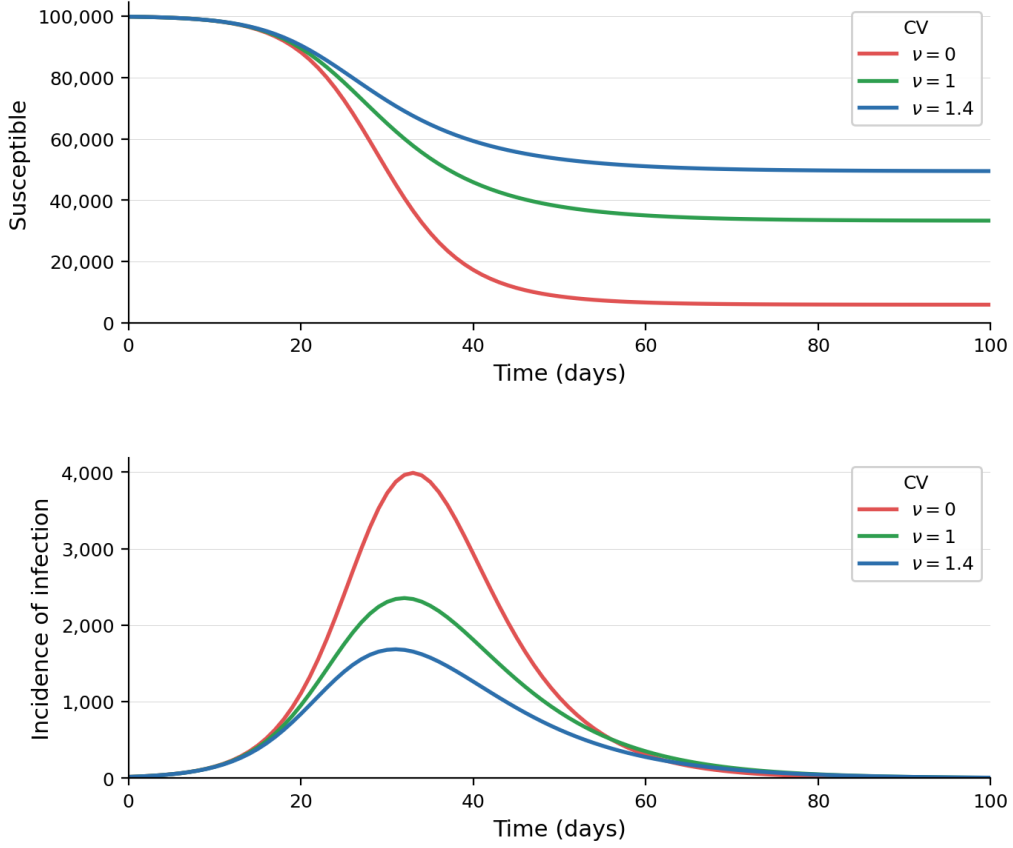


Figure 1.5: Effect of heterogeneity on epidemic dynamics for three values of the coefficient of variation ($\nu = 0, 1, 1.4$). *Top:* susceptible pool over time. Higher ν leaves substantially more individuals uninfected at the epidemic's end. *Bottom:* corresponding daily incidence (δE), showing that heterogeneity lowers and broadens the epidemic peak. The homogeneous model ($\nu = 0$, red) produces the largest and sharpest peak. *Parameters:* $N = 100,000$; $\mathcal{R}_0 = 3.0$; latent period $1/\delta = 5.5$ days; infectious period $1/\gamma = 4$ days; relative infectiousness $\rho = 0.5$; initial seeds $E_0 = 100$, $I_0 = 40$. No NPIs applied.

Here β is the baseline transmission parameter in the absence of intervention, while $c(t)$ is the proportion of baseline effective contact retained at time t . Thus $c(t) = 1$ means no reduction in transmission, whereas $c(t) < 1$ represents reduced effective contact.

When the reduced heterogeneous SEIR model is combined with this time-varying transmission coefficient, the susceptible equation becomes

$$\frac{dS}{dt} = -\beta c(t) (\rho E(t) + I(t)) \left(\frac{S(t)}{N} \right)^{1+\nu^2}. \quad (1.97)$$

The remaining equations are

$$\frac{dE}{dt} = \beta c(t) (\rho E(t) + I(t)) \left(\frac{S(t)}{N} \right)^{1+\nu^2} - \delta E(t), \quad (1.98)$$

$$\frac{dI}{dt} = \delta E(t) - \gamma I(t), \quad (1.99)$$

$$\frac{dR}{dt} = \gamma I(t). \quad (1.100)$$

This is the main model form underlying the simulation and inference chapters of the thesis.

1.7.1 Autonomous and non-autonomous epidemic systems

The distinction between heterogeneous susceptibility and NPIs is both epidemiological and mathematical. When β , δ , γ , and ρ are constant and $c(t) \equiv 1$, the heterogeneous SEIR model is autonomous: its right-hand side depends on the current state (S, E, I, R), but not explicitly on calendar time. Epidemic slowdown then arises endogenously through susceptible depletion and the preferential removal of highly susceptible individuals.

Replacing β by $\beta c(t)$ makes the model non-autonomous because transmission now depends explicitly on time through the prescribed contact multiplier $c(t)$. NPIs therefore reduce transmission exogenously through time-varying contact behaviour, whereas heterogeneity acts through changes in the composition of the susceptible population.

With both mechanisms present,

$$\mathcal{R}_{\text{eff}}(t) = \mathcal{R}_0 c(t) \left(\frac{S(t)}{N} \right)^{1+\nu^2}. \quad (1.101)$$

The factor $c(t)$ represents time-varying contact reduction, while $\left(\frac{S(t)}{N} \right)^{1+\nu^2}$ represents susceptible depletion and selective depletion under Gamma-distributed susceptibility. Although the two mechanisms enter the model differently, both can produce similar reductions in incidence, giving rise to the identifiability problem studied in later chapters. This is briefly shown in Chapter 2, Section 2.2, Figure 2.1.

1.7.2 Parametric NPI functional forms

The simplest specification treats intervention as instantaneous, so that contact drops to its reduced level on the day the measure is imposed:

$$c_{\text{step}}(t) = \begin{cases} 1, & t \leq t_0, \\ c^*, & t > t_0, \end{cases} \quad (1.102)$$

where t_0 is the onset time and $c^* \in (0, 1]$ the sustained contact level retained afterwards; $c^* = 1$ is no intervention, and smaller c^* a stronger reduction. This is the form used when transmission is switched at announced policy dates [59, 60] or at inferred change points [61], and McCarthy and Dobrovolny [62] identify it as the most common assumption in this literature. Its weakness is that behaviour is rarely abrupt and policy takes time to propagate through a population.

The remaining forms smooth that transition, and each returns (1.102) as a limiting case. The baseline profile used in the simulation chapters of this thesis is a piecewise-linear ramp,

$$c_{\text{lin}}(t) = \begin{cases} 1, & t \leq t_0, \\ 1 - (1 - c^*) \frac{t - t_0}{t_1 - t_0}, & t_0 < t \leq t_1, \\ c^*, & t > t_1, \end{cases} \quad (1.103)$$

in which t_0 is the onset of contact change, t_1 the time at which the sustained level is reached, and $t_1 - t_0$ the transition duration. The two timing parameters admit a two-event reading, with t_0 marking voluntary behavioural change and t_1 the imposition of formal measures. Deforche et al. [63] analysed first-wave cases and deaths across 35 Western countries and found that behavioural change preceding lockdown, prompted by news of the epidemic in Lombardy, made a substantial contribution to the observed reduction in transmission. Ramps of this shape are used by Gomes et al. [4], whose $c(t)$ for England and Scotland combines a linear reduction before the first lockdown with a linear return towards baseline afterwards, parameterised by the estimated durations t_0 and t_2 . A second alternative is an exponential transition,

$$c_{\text{exp}}(t) = \begin{cases} 1, & t \leq t_0, \\ c^* + (1 - c^*) \exp\{-\kappa(t - t_0)\}, & t > t_0, \end{cases} \quad (1.104)$$

where $\kappa > 0$ sets the decay rate, so contact falls fastest at onset and approaches c^* asymptotically. A third is a logistic transition,

$$c_{\text{log}}(t) = 1 - \frac{1 - c^*}{1 + \exp\{-k(t - t_m)\}}, \quad (1.105)$$

with midpoint t_m and steepness $k > 0$, appropriate when the strongest behavioural change is assumed to occur around a central time rather than immediately at t_0 . Both smooth forms are fitted by McCarthy and Dobrovolny [62]. The step form is the common limit of all three: of (1.103) as $t_1 \rightarrow t_0$, of (1.104) as $\kappa \rightarrow \infty$, and of (1.105) as $k \rightarrow \infty$. Figure 1.6 compares the linear, exponential, and logistic profiles after matching them to the same onset, transition window, and sustained contact level (See Appendix A.2 for the details of matching the transition windows). Although the profiles imply similar overall reductions in transmission, they differ in the timing and smoothness of the decline, which can affect inference about intervention onset, strength, and the contribution of susceptibility heterogeneity.

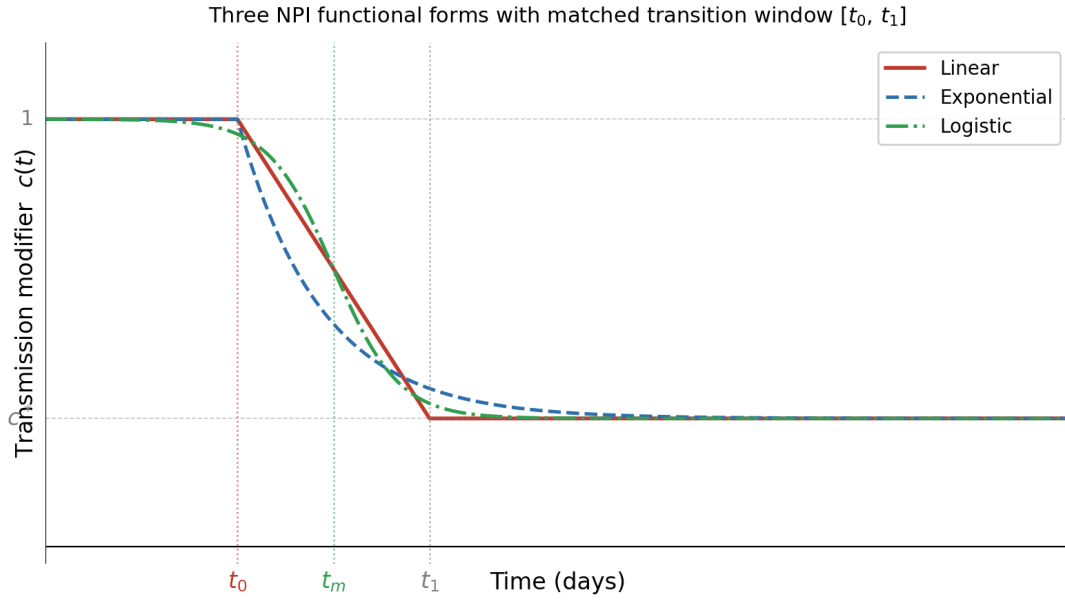


Figure 1.6: The three gradual NPI functional forms with a matched transition window $[t_0, t_1]$. The linear profile decreases at a constant rate between t_0 and t_1 . The exponential profile begins declining at t_0 and approaches the sustained contact level c^* asymptotically. The logistic profile has its steepest decline around the midpoint t_m . Each tends to the step profile (1.102) as its transition width shrinks to zero.

The choice among these forms is consequential rather than cosmetic. If the true behavioural response changes gradually but the fitted model assumes an abrupt step, the fitted intervention effect may be biased; conversely, if the fitted profile is too flexible, several different transitions may explain the same epidemic curve almost equally well. McCarthy and Dobrovolny [62] make the point empirically: they fitted all four forms to cumulative case counts from 51 US regions and selected between them by BIC, and the exponential was preferred in 24 regions and the logistic in 18, against 5 for the linear ramp and 4 for the instantaneous step. They further report that the estimated onset of contact reduction generally fell later than the declared lockdown date, with the discrepancy largest under the step model, and that in some regions it fell earlier, which they read as evidence of self-isolation preceding government action. The specification of $c(t)$ is therefore part of the inferential model, not a presentational choice made after the modelling is settled.

That last finding also points to a difficulty that no choice of functional form removes. Observed case counts or deaths do not respond immediately to a change in transmission: infections must first occur, pass through incubation and reporting delays, and, in mortality analyses, pass through the infection-to-death delay. Consequently, the timing of a change in the observed epidemic curve can lag behind the timing of the underlying behavioural or policy change. This makes parameters such as t_0 , t_1 , t_m , and c^* partly dependent on the assumed observation model and

delay distribution, and not only on the transmission model itself [61].

A further difficulty is whether the parameters of a given profile are identifiable at all. Soltesz et al. [64] argue that the effects of individual step-modelled NPIs are not separately identifiable because the measures were implemented close together in time, and caution that a large estimated effect is not evidence that the corresponding parameter is identified. In reply, Flaxman et al. [65] accept that the relative effects cannot be disentangled for a single country considered alone, attributing this partly to the spread of the infection-to-death delay, but maintain that joint analysis across countries recovers an identifiable signal. That exchange prefigures the argument of this thesis: the specification of $c(t)$ belongs to the inferential problem, and information pooled across epidemics is one route out of it (Section 1.8).

1.7.3 Stringency-driven contact profiles

A different approach is to use an external policy measure rather than estimating the full shape of $c(t)$ from the epidemic data. In Chapter 6, one NPI specification uses the Oxford COVID-19 Government Response Tracker stringency index as an externally observed summary of government restrictions. Let $SI_{j,t} \in [0, 100]$ denote the stringency index for country j on day t . A simple stringency-driven contact multiplier is

$$c_{j,t} = c^* + (1 - c^*) \left(1 - \frac{SI_{j,t}}{100} \right). \quad (1.106)$$

This mapping gives $c_{j,t} = 1$ when the stringency index is zero, and $c_{j,t} = c^*$ when the stringency index is 100. Thus c^* again represents the minimum sustained contact level under maximal restrictions.

The stringency-driven profile has a different interpretation from the parametric profiles above. In the piecewise-linear, exponential, and logistic models (Section 1.7.2), the timing and shape of contact reduction are inferred through parameters such as t_0 , t_1 , t_m , k , or κ . In the stringency-driven model, much of that time variation is instead supplied externally by an observed policy series: the Oxford COVID-19 Government Response Tracker provides daily policy indicators and composite indices, including the Stringency Index, for cross-national comparison of government responses [39]. This reduces the number of timing parameters that must be learned from mortality or incidence data, and with them one route by which a fitted profile can absorb other features of the curve. The cost is a different assumption, namely that the policy index is a suitable proxy for transmission-relevant behavioural change. If behaviour changed before formal policy, if adherence varied between countries, or if important behavioural changes were

not captured by the index, the stringency-driven $c(t)$ is misspecified in its own way.

One limitation is common to every specification in this section, parametric and stringency-driven alike: a scalar contact multiplier cannot represent all the ways in which interventions change behaviour. Real interventions may alter where contacts occur, which individuals continue to mix, and how contacts are distributed across households, workplaces, schools, and social settings. Contact surveys during the COVID-19 pandemic showed that interventions changed contact patterns and not merely the overall level of contact [66,67]. A scalar $c(t)$ is thus a useful low-dimensional summary, but it may absorb several different mechanisms: average contact reduction, changes in the distribution of exposure, and changes in reporting or ascertainment.

These issues matter directly for this thesis because susceptibility heterogeneity can also generate epidemic slowdown. If heterogeneity is omitted, an inference procedure may attribute intrinsic selective depletion to external contact reduction. If the NPI profile is misspecified, the fitted model may compensate by changing the estimated heterogeneity parameter ν , the onset time t_0 , or the sustained contact level c^* . The same observed decline can therefore support different explanations unless the model, data and likelihood contain enough information to separate the mechanisms.

Table 1.4 summarises the inference-centred reason for bringing the heterogeneity and NPI literatures together. First, heterogeneity changes epidemic thresholds, final sizes and the interpretation of epidemic slowdown. Second, estimated NPI effects depend on the assumed transmission modifier, delay model, and statistical structure. Third, when both mechanisms are present, the epidemic curve may identify a composite reduction in transmission better than it identifies the biological and behavioural mechanisms separately. This motivates the combined heterogeneity–NPI identifiability problem introduced in the next section.

1.8 The Combined Heterogeneity–NPI Identifiability Problem

The preceding sections reviewed two mechanisms that can reduce epidemic growth. The first is intrinsic: in a heterogeneous population, individuals with higher susceptibility tend to be infected earlier, so the remaining susceptible population becomes progressively less susceptible on average. The second is extrinsic: non-pharmaceutical interventions and behavioural changes reduce transmission through the time-varying contact multiplier $c(t)$. The central inferential question of this thesis is how far these two mechanisms are practically identifiable from epidemic time-series data when both are included in the same model.

For the Gamma-susceptibility SEIR model with NPI scaling, the effective reproduction

Table 1.4: Key results linking heterogeneity, NPIs, and inferential uncertainty

Theme	Representative papers	Main result, as used in this thesis	Implication for inference
Heterogeneity lowers herd-immunity thresholds and final sizes relative to homogeneous baselines	Britton et al. [37]; Gomes et al. [4]; Montalbán et al. [44]	Allowing individual variation in susceptibility or exposure can substantially reduce inferred herd-immunity thresholds and epidemic final sizes under the model assumptions.	Early-wave trajectories alone may not pin down heterogeneity strength: different combinations of heterogeneity and time-varying transmission can fit similarly well.
NPIs as step changes or change points	Flaxman et al. [59]; Dehning et al. [61]	Transmission is commonly represented through a time-varying modifier, either aligned with intervention dates or inferred through change points from time series.	If heterogeneity is omitted, intrinsic slowdown can be absorbed into inferred intervention effects; reporting and mortality delays also blur intervention timing signals.
Model uncertainty in NPI effects	Chin et al. [68]; Müller et al. [69]	Plausible competing models can yield materially different estimates of NPI effect sizes and uncertainty.	NPI effect estimates can be model-dependent, reinforcing the need to treat $c(t)$ specification as part of the inferential problem.
Functional form of NPI matters	McCarthy and Dobrovolny [62]	Comparisons of alternative NPI shapes, such as exponential and logistic decay functions, can change fit quality and inferred timing or strength.	NPI parameterisation can interact with heterogeneity during fitting; misspecified $c(t)$ can worsen confounding between ν and c^* .
Behavioural-response parameters are hard to estimate early	Osi and Ghaffarzadegan [70]	Endogenous risk-response models introduce additional behavioural parameters that can be weakly identified when behaviour and policy shift quickly.	Behavioural dynamics add another layer of practical identifiability difficulty during first-wave inference windows.
Covariation of susceptibility and transmissibility	Harris et al. [58]	Correlation between susceptibility and transmissibility can change epidemic speed and size relative to independent-variation assumptions.	Misspecifying the structure of heterogeneity can shift inferred intervention strength and/or heterogeneity magnitude.

number can be written as

$$\mathcal{R}_{\text{eff}}(t) = \mathcal{R}_0 c(t) \left(\frac{\mathbf{S}(t)}{N} \right)^{1+\nu^2}. \quad (1.107)$$

This expression shows the source of the inferential overlap. The factor $c(t)$ reduces transmission externally through changes in contact behaviour. The factor $\{\mathbf{S}(t)/N\}^{1+\nu^2}$ reduces transmission internally through susceptible depletion and selective depletion under Gamma-distributed susceptibility. Therefore, both a smaller contact multiplier and a larger coefficient of variation can flatten an epidemic curve.

It is useful to rewrite (1.107) as

$$\mathcal{R}_{\text{eff}}(t) = \mathcal{R}_0 \left(\frac{\mathbf{S}(t)}{N} \right) \underbrace{c(t) \left(\frac{\mathbf{S}(t)}{N} \right)^{\nu^2}}_{\text{combined contact-heterogeneity factor}}. \quad (1.108)$$

During the intervention plateau, $c(t) \approx c^*$. Over a short part of a single epidemic wave, the susceptible fraction may also vary within a limited range. If $\mathbf{S}(t)/N$ is approximated locally by a representative value $\mathbf{S}^*/N$, then the data may mainly inform the composite quantity

$$c_{\text{eff}} \approx c^* \left(\frac{\mathbf{S}^*}{N} \right)^{\nu^2}. \quad (1.109)$$

Since $0 < \mathbf{S}^*/N < 1$, increasing ν reduces $\{\mathbf{S}^*/N\}^{\nu^2}$. This reduction can be partly offset by increasing c^* , that is, by inferring a weaker contact reduction. Conversely, a smaller ν can be partly offset by a smaller c^* , corresponding to a stronger inferred intervention. This gives the intuitive direction of the ν - c^* compensation studied in the later chapters.

Figure 1.7 illustrates this local compensation for a representative susceptible fraction. The curve shows combinations of (ν, c^*) that keep c_{eff} approximately constant. Moving along the curve therefore leaves the combined contact-heterogeneity factor nearly unchanged, even though the individual values of ν and c^* differ.

The same point can be expressed on the logarithmic scale:

$$\log \mathcal{R}_{\text{eff}}(t) = \log \mathcal{R}_0 + \log c(t) + (1 + \nu^2) \log \left(\frac{\mathbf{S}(t)}{N} \right). \quad (1.110)$$

This form makes clear that changes in $\log c(t)$ and changes in $\nu^2 \log\{\mathbf{S}(t)/N\}$ can act in opposite directions while leaving the fitted epidemic trajectory similar. The compensation is not expected to be exact over the whole epidemic, because $\mathbf{S}(t)$, $\mathbf{E}(t)$, and $\mathbf{I}(t)$ evolve through time. However, it can still produce a shallow likelihood ridge when the available data are finite and noisy.

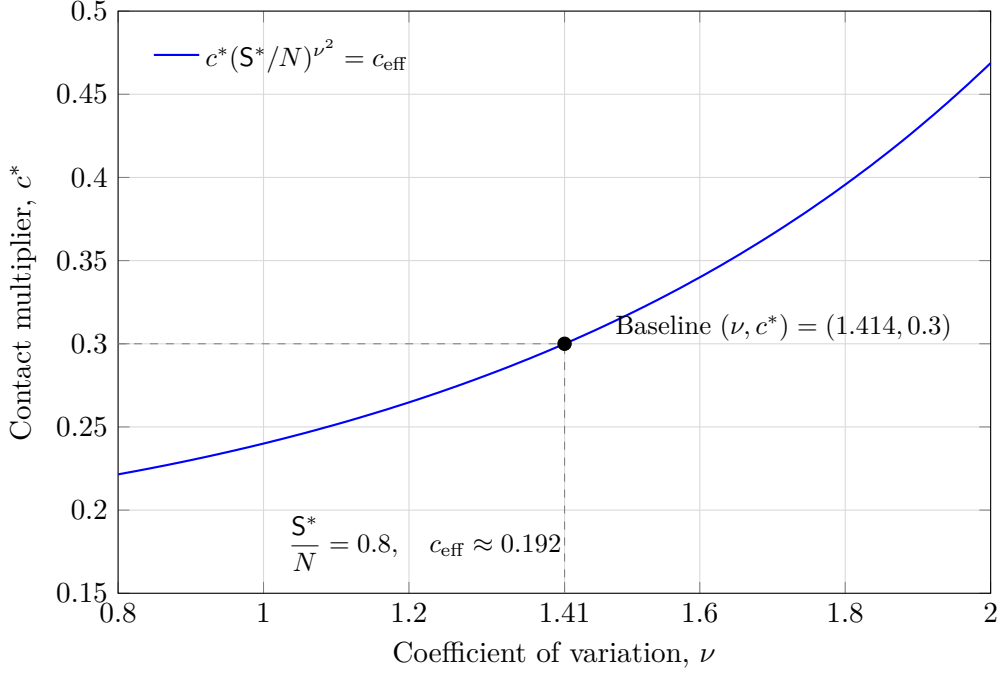


Figure 1.7: Illustration of local compensation between the contact multiplier c^* and the susceptibility coefficient of variation ν under the approximation $c_{\text{eff}} \approx c^*(S^*/N)\nu^2$. The curve shows combinations of (ν, c^*) that keep the composite quantity c_{eff} constant at the value determined by the baseline point $(\nu, c^*) = (1.414, 0.3)$, using the representative susceptible fraction $S^*/N = 0.8$. Along this curve, an increase in ν must be accompanied by an increase in c^* to preserve the same approximate transmission-reducing effect.

This is why the four main inferential targets in this thesis are

$$\theta = (\mathcal{R}_0, \nu, t_0, c^*). \quad (1.111)$$

Each parameter controls a different scientifically important part of the epidemic interpretation. The parameter $\mathcal{R}_0$ controls baseline transmission before depletion and before intervention. The parameter ν controls the strength of susceptibility heterogeneity and therefore the degree of selective depletion. The parameter t_0 controls when contact reduction begins. The parameter c^* controls the sustained change to the level of contact after the intervention has been established. Estimating these four quantities jointly is therefore the simplest way to ask whether epidemic slow-down should be attributed mainly to baseline transmissibility, intrinsic heterogeneity, intervention timing, or intervention strength.

1.9 Research Gap and Thesis Contribution

The literature has made substantial progress along two related directions. On one hand, there is a clear development from homogeneous SEIR models to frameworks that account for individual

variation in susceptibility, including Gamma-distributed frailty models. On the other hand, there is a large literature on representing NPIs through time-varying transmission modifiers and estimating their effects from epidemic data.

What is still missing is a robust and systematic approach for separating these two mechanisms when fitting to standard single-epidemic time series. When heterogeneity and NPIs are both included, they can explain the same broad patterns in observed incidence or deaths, so different combinations of ν and $c(t)$ can fit similarly well. This may create a practical identifiability problem that undermines how reliably we can interpret parameter estimates as either intrinsic population heterogeneity or the effect of interventions.

This thesis responds to that gap. We will show using simple input-output method in Chapter 2, that under idealised conditions that ν and $c(t)$ are structurally identifiable and then investigate their practical identifiability (along with other transmission potentials). We will keep the heterogeneous SEIR framework as a foundation, but with focus on its main inferential limitation. The later Chapters study identifiability directly and then test whether using information from joint-epidemics inference can weaken the compensation between heterogeneity and interventions. This idea is consistent with existing multi-location fitting strategies that share heterogeneity and intervention profiles across England and Scotland [4], and it provides the bridge from the reviewed literature to the original contributions of the thesis.

1.10 Aim and Objectives

The main aim of this thesis is to develop and evaluate a framework for simultaneous inference of heterogeneity in susceptibility and impact of NPIs on epidemic trajectories using models that include both.

More specifically, the thesis studies SEIR-type models in which susceptibility heterogeneity is represented by a Gamma distribution, or by related discretised susceptibility distributions, and NPI effects are represented through a time-varying contact multiplier $c(t)$.

The four main inferential targets are

$$(\mathcal{R}_0, \nu, t_0, c^*),$$

namely the basic reproduction number, the coefficient of variation in susceptibility, the onset time of contact reduction, and the sustained post-intervention contact level. These quantities are targeted because they separate the main scientific mechanisms in the model, for the questions

the thesis seeks to investigate, as explained in section 1.8.

The objectives are:

1. to assess, using single-epidemic simulation studies, the practical identifiability of $(\mathcal{R}_0, \nu, t_0, c^*)$ and to characterise the dominant compensation mechanisms;
2. to develop and evaluate a multi-epidemic inference framework, and to assess how far joint analysis across epidemics can reduce parameter confounding and tighten uncertainty;
3. to develop discretisation methods for continuous susceptibility distributions, so that simulation studies can use structured heterogeneous-susceptibility models as well as reduced-form models;
4. to study robustness to modelling assumptions, including susceptibility-distribution misspecification and NPI functional-form choices, and to examine the consequences for parameter recovery and forecasting;
5. to apply the resulting framework to COVID-19 mortality data from England and Scotland and assess what can be learned empirically about heterogeneity and intervention strength.

1.11 Thesis Outline

The thesis is organised as follows. Chapter 2 studies practical identifiability using a single epidemic. Chapter 3 extends the analysis to a joint-epidemic framework. Chapter 4 develops discretisation methods for continuous susceptibility distributions. Chapter 5 examines robustness to distributional misspecification. Chapter 6 applies the framework to COVID-19 mortality data from England and Scotland. Chapter 7 concludes the thesis by summarising the main findings, discussing limitations, and outlining directions for future work.

Chapter 2

Heterogeneity in Susceptibility and NPIs: Parameter Identifiability

2.1 Introduction

In the literature review of the previous chapter, we outlined the framework and research questions for simultaneous inference of **heterogeneity in susceptibility** (ν) and **non-pharmaceutical interventions (NPIs)** (c^*), which flatten epidemic curves via distinct mechanisms. When both effects are present, separating their contributions is non-trivial and can lead to parameter non-identifiability.

In this chapter we examine **practical identifiability** (as defined in Section 2.3) of a heterogeneous SEIR model with NPIs, using synthetic epidemic data. Specifically, we investigate how well a *single epidemic time series* can inform four key parameters of interest in the reduced heterogeneous-susceptibility model (2.1): the basic reproduction number ($\mathcal{R}_0$), the coefficient of variation (CV) of susceptibility (ν), the onset time of behavioural change (t_0), and the NPI strength (c^* , the sustained contact multiplier). Our goal is to understand how these parameters interact in the likelihood surface, and to assess the consequences for parameter estimation, uncertainty quantification, and model comparison.

Objectives of the chapter

Our aim in this chapter is to study the performance and limitations of the heterogeneous-susceptibility SEIR model (2.1) for *single* epidemic trajectories. Our objectives are:

1. **Clarify the structural context.** Provide a brief input–output argument that, under

idealised assumptions, the four key parameters $(\mathcal{R}_0, \nu, t_0, c^*)$ are structurally identifiable, while emphasising that the focus of this chapter is on *practical* identifiability.

2. **Quantify mis-specification bias.** Assess bias and precision when simulating from a homogeneous or heterogeneous susceptibility model and fitting the alternative model, with and without NPIs.
3. **Characterise practical identifiability in a single epidemic.** Analyse the geometry of the likelihood surface for a single epidemic, with particular attention to the ν – c^* correlation and compensating effects. We examine how this relationship manifests across different NPI profiles and estimation methods.
4. **Evaluate forecasting implications.** Compare homogeneous and heterogeneous fits in terms of their out-of-sample forecasts for a two-wave epidemic trajectory.

We organise the chapter as follows. First, we restate the reduced heterogeneous-susceptibility transmission model and the baseline NPI functional form used in the simulation study (Section 2.2). In Section 2.3, we distinguish structural and practical identifiability. We then give an input–output argument for the structural context of the model (Section 2.3.1) and develop the mathematical intuition for the source of the ν – c^* compensation ridge. In Section 2.4, we describe the baseline practical-identifiability analysis, including the simulation design (Section 2.4.1) and the maximum-likelihood estimation procedure (Section 2.5). The baseline results are then presented in Section 2.6, first for homogeneous data-generating scenarios (Sections 2.6.1–2.6.2) and then for heterogeneous data-generating scenarios (Sections 2.6.3–2.6.4). After the deterministic baseline analysis, Section 2.7 examines robustness to stochastic process noise using τ -leaping simulations across different population sizes. Section 2.8 then studies the forecasting implications of the fitted homogeneous and heterogeneous models. Finally, Section 2.9 summarises the main findings of the chapter.

Two appendices provide supplementary robustness checks and sensitivity analyses. Appendix B investigates whether the ν – c^* correlation persists across different estimation methods and NPI functional forms. Appendix C shows that, although the strength of control can affect the sharpness of the ridge, the basic single-epidemic difficulty remains present across the tested range $c^* \in \{0.20, 0.30, 0.40\}$.

2.2 Mathematical Model and NPIs

The heterogeneous-susceptibility SEIR model and the baseline piecewise-linear NPI profile were introduced in Chapter 1, section 1.6.3, equations (1.78)–(1.81). Since this is the first inferential chapter, we restate the reduced model and the baseline intervention parameterisation used in the simulation study.

2.2.1 Transmission model

Throughout this chapter, we use the reduced heterogeneous-susceptibility SEIR model of Gomes et al. [4], in which heterogeneity is summarised by the coefficient of variation ν . The homogeneous model is recovered by setting $\nu = 0$:

$$\begin{aligned}\frac{dS}{dt} &= -\beta(\rho E + I) \left(\frac{S}{N}\right)^{1+\nu^2}, \\ \frac{dE}{dt} &= \beta(\rho E + I) \left(\frac{S}{N}\right)^{1+\nu^2} - \delta E, \\ \frac{dI}{dt} &= \delta E - \gamma I, \\ \frac{dR}{dt} &= \gamma I.\end{aligned}\tag{2.1}$$

The basic reproduction number is

$$\mathcal{R}_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right).\tag{2.2}$$

The values of δ, γ, ρ are given in Table 2.1. The initial conditions are fixed at

$$I_0 = 40, \quad E_0 = 2.5I_0 = 100, \quad S_0 = N - E_0 - I_{tot0}, \quad R(0) = 0.$$

The ratio $E_0/I_0 \approx 2.5$ is motivated by the early-growth linearisation in Appendix A.1.3; the corresponding numerical values are listed in Table 2.1.

2.2.2 Baseline NPI parameterisation

To represent non-pharmaceutical interventions and behavioural change, we scale transmission by a time-varying contact multiplier $c(t) \in (0, 1]$, so that $\beta(t) = \beta c(t)$. The baseline specification in this chapter is the same piecewise-linear profile (because it gives a direct epidemiological interpretation to the two main timing quantities in the simulation design) for $c(t)$ introduced in

Chapter 1, Section 1.7 illustrated in Figure 2.1:

$$c(t) = \begin{cases} 1, & t \leq t_0, \\ 1 - (1 - c^*) \frac{t - t_0}{t_1 - t_0}, & t_0 < t \leq t_1, \\ c^*, & t > t_1, \end{cases} \quad (2.3)$$

where t_0 is the onset of behavioural change, t_1 marks the start of maximal contact reduction, and c^* is the sustained contact level. In the baseline simulation study, t_1 is fixed (represents the time at which the strongest contact reduction is reached, which can be aligned with a lockdown or formal policy date), while t_0 and c^* are estimated jointly with the epidemiological parameters.

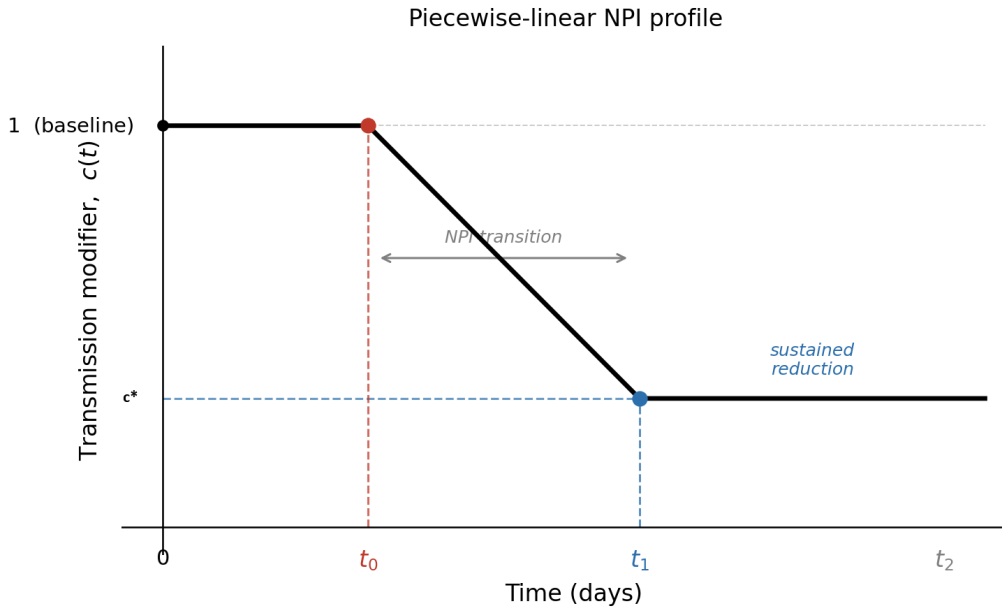


Figure 2.1: Piecewise-linear NPI profile $c(t)$. Transmission is unchanged at baseline ($c(t) = 1$) until time t_0 , when individuals begin voluntarily reducing contacts in response to news of the emerging epidemic, before any official government mandate. From t_0 to t_1 , $c(t)$ decreases linearly as awareness grows and behavioural change spreads through the population. At t_1 , government-mandated control measures (such as lockdown) come into force and transmission is held at the sustained reduced level c for $t > t_1$.

Figure 2.2 compares epidemic trajectories under heterogeneity in susceptibility and under non-pharmaceutical interventions (NPIs), where NPIs are represented by the piecewise linear contact reduction in (2.3). Both mechanisms reduce peak incidence and total cases relative to the homogeneous model without interventions, but they do so in different ways: heterogeneity mainly lowers and broadens the peak through selective depletion of highly susceptible individuals, whereas NPIs reduce transmission through an explicit time-varying reduction in contacts, which can visibly change the timing and skewness of the epidemic curve.

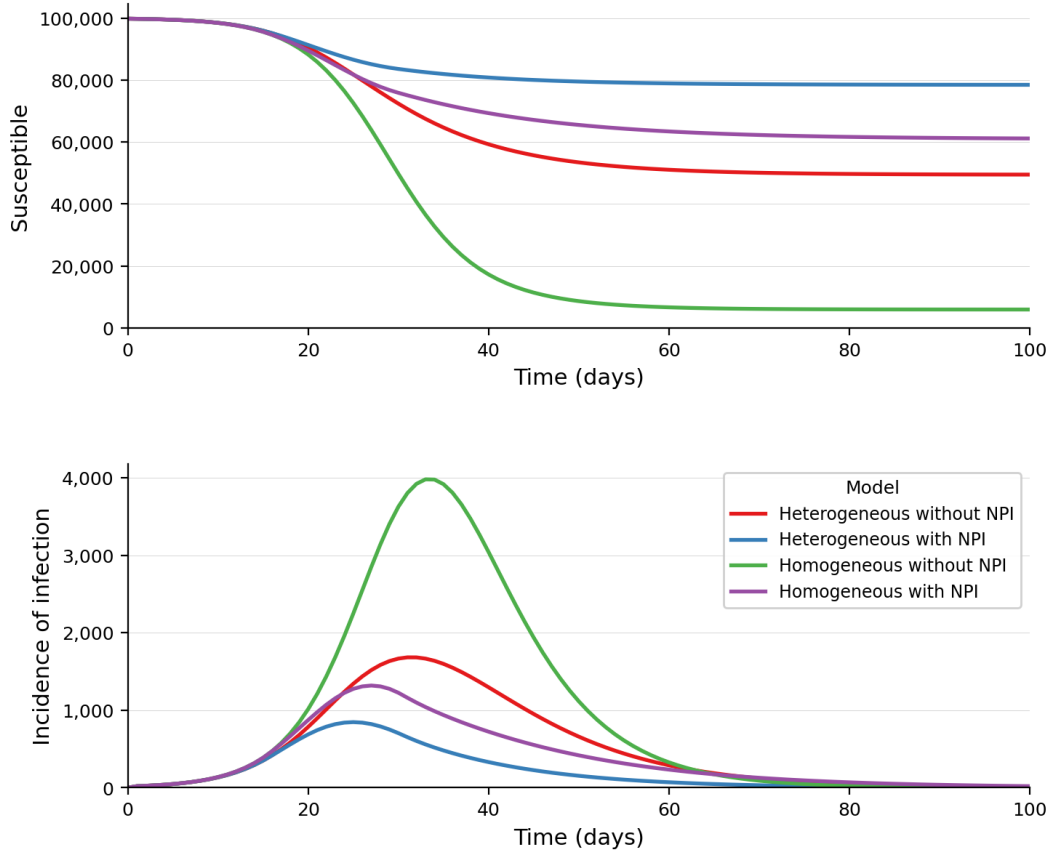


Figure 2.2: Epidemic trajectories under four combinations of heterogeneity and NPIs (equation (2.3)) for $\mathcal{R}_0 = 3$ and a population of 100,000. NPI parameters: $t_0 = 15$ days (onset of voluntary behavioural change), $t_1 = 20$ days (start of official lockdown), sustained contact level $c^* = 0.3$; heterogeneous curves use $\nu = 1.414$. *Top:* susceptible pool over time. *Bottom:* daily incidence of infection. Both heterogeneity and NPIs reduce peak incidence relative to the homogeneous baseline without intervention (green), but through different mechanisms: heterogeneity lowers and broadens the peak endogenously via selective depletion of high-susceptibility individuals, while NPIs reduce transmission exogenously through $c(t)$.

Visually, these effects can look similar because both “flatten” the trajectory. The key question is therefore whether, from a single observed time series, we can reliably separate changes driven by intrinsic heterogeneity from changes driven by interventions. This motivates the discussion of structural and practical identifiability in Section 2.3.

2.2.3 Choice of epidemiological parameters

The parameter values used in this thesis are motivated by empirical estimates from the COVID-19 pandemic, particularly in the United Kingdom. Early studies of SARS-CoV-2 transmission consistently estimated the basic reproduction number in the range $\mathcal{R}_0 \approx 2.5$ – 3.5 , with central estimates close to 3.0 [71–73]. Accordingly, we adopt $\mathcal{R}_0 = 3.0$ as a representative baseline value. Empirical and modelling studies suggest that the coefficient of variation ν in susceptibility lies in the range 1.0–2.0, with values around 1.4–1.5 being consistent with UK-based analyses [4].

Table 2.1: Baseline simulation parameters motivated by empirical estimates from the COVID-19 pandemic. In particular, $\mathcal{R}_0 \approx 3.0$ reflects early transmission estimates in the UK, while the coefficient of variation $\nu = 1.414$ is consistent with reported levels of heterogeneity in susceptibility. Other epidemiological parameters are taken from [4].

Parameter	Description	Value	Source / notes
N	Total population	100,000	Assumed
$1/\delta$	Mean exposed period before entry into I	5.5 days	[4]
$1/\gamma$	Mean fully infectious period	4 days	[4]
ρ	Relative infectiousness in E	0.5	[4]
$\mathcal{R}_0$	Basic reproduction number	3.0	Fixed
ν	Coefficient of variation in susceptibility	1.414	Fixed
c^*	Sustained contact multiplier (NPI floor)	0.3	Fixed
t_0	Behavioural change onset time	15 days	Fixed
t_1	Time to beginning of lockdown	20 days	Fixed
I_0	Initial fully infectious individuals	40	Assumed baseline seed
E_0	Initial exposed individuals	100	Set as $E_0 = 2.5I_0$; see Appendix A.1.3
S_0	Initial susceptible individuals	$N - E_0 - I_0 =$ 99,860	Implied by closed population
$R(0)$	Initial removed individuals	0	Assumed

Motivated by these findings, we set $\nu = \sqrt{2}$ in the baseline simulations, corresponding to variance 2 for a unit-mean susceptibility distribution. Sensitivity analyses conducted in Appendix A.4 show that the qualitative conclusions remain unchanged when $\mathcal{R}_0$ varies within the range 2.5–3.5.

2.3 Structural and Practical Identifiability

We briefly distinguish between *structural* and *practical* identifiability, following the classical control-theoretic definition of structural identifiability (see e.g Bellman and Åström [74], Villaverde et al. [75], Wieland et al. [76]).

Structural identifiability. Consider a general dynamical system

$$\dot{x}(t) = f(x(t), \theta, u(t)), \quad y(t) = g(x(t), \theta, u(t)), \quad (2.4)$$

where $x(t)$ is the state vector, θ is the parameter vector, $u(t)$ denotes any inputs, and $y(t)$ is the observed output. Under ideal conditions (noise-free continuous observation of y over a sufficiently long time interval, known inputs u and known initial conditions), *structural identifiability* asks

whether the mapping

$$\theta \longmapsto y(\cdot; \theta)$$

is one-to-one (at least locally), Villaverde et al. [75]. A parameter θ_i is structurally identifiable if

$$y(t; \theta) = y(t; \theta') \text{ for all } t \implies \theta_i = \theta'_i$$

for all admissible parameter vectors θ, θ' . If every component of θ satisfies this property, the model is said to be structurally identifiable. This is a property of the model structure (2.4) (the equations and output definition), not of any particular dataset [75–77].

Practical identifiability. In applications we never have access to idealised, noise-free, continuous data. *Practical identifiability* instead concerns whether, for a given experimental design and noise level, the available finite and noisy observations contain enough information to estimate the parameters with useful precision (see Raue et al. [78] and Wieland et al. [76] for detailed discussions). Even if the model (2.4) is structurally identifiable in the sense above, parameters can be *practically* non-identifiable for several reasons that are well documented in the literature, for example:

- **Insufficient or poorly located data.** If the observation window is too short, the sampling times are sparse, or key dynamical features (such as the epidemic peak or decay phase) are not observed, then the likelihood surface can be very flat in some parameter directions, leading to broad profile likelihoods and large uncertainty intervals [76–78].
- **Measurement noise.** Large observational noise can mask the changes in $y(t)$ induced by varying a parameter, again producing very shallow likelihood profiles and practically uninformative estimates [76, 78].
- **Compensating parameter effects.** Different parameters may influence the output in similar or compensatory ways, so that their effects cannot be cleanly separated from the data. This manifests as long, narrow ridges or valleys in the likelihood or cost function, a phenomenon often described as “sloppiness” in nonlinear dynamical models [76, 78, 79].

In such situations, point estimates of θ can be highly unstable, and apparently good fits may correspond to wide regions of parameter space [76, 78]. The central aim of this chapter is therefore to study the *practical* identifiability of our heterogeneous SEIR model when extrinsic NPIs are present, under realistic data conditions, rather than to provide an exhaustive formal analysis of

its structural identifiability.

2.3.1 Structural Identifiability: Input–Output Reasoning

We use a simple input–output argument, based on the differential-algebra approach formalised by Ollivier [80], to examine when different parameter sets can produce identical observable outputs. The argument is informal but clarifies why, under idealised conditions, the parameter vector $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$ is structurally distinguishable in the reduced heterogeneous SEIR model (2.1). As before, S, E, I denote susceptible, exposed and infectious counts, the output is the incidence $y(t) := \delta E(t)$, and

$$\beta(t) := \frac{\mathcal{R}_0 c(t)}{\rho/\delta + 1/\gamma}, \quad c(t) := g(t - t_0; c^*), \quad (2.5)$$

with g a fixed known shape, $g(t) = 1$ for $t \ll 0$ and $g(t) \rightarrow c^*$ as $t \rightarrow \infty$. Throughout we assume ρ, δ, γ, N and g known, and the initial conditions (S_0, E_0, I_0) fixed and common across compared parameter sets.

Lemma 2.3.1 (Structural identifiability under idealised observation). *Consider the reduced heterogeneous SEIR model (2.1) with output $y(t) = \delta E(t)$ observed continuously and without noise on an interval $[0, T]$ spanning a non-trivial epidemic, under the standing assumptions above. Let*

$$\theta_1 = (\mathcal{R}_{0,1}, \nu_1, t_{0,1}, c_1), \quad \theta_2 = (\mathcal{R}_{0,2}, \nu_2, t_{0,2}, c_2)$$

be two admissible parameter vectors producing identical outputs,

$$y_1(t) = y_2(t) \quad \text{for all } t \in [0, T].$$

Then $\theta_1 = \theta_2$. That is, $(\mathcal{R}_0, \nu, t_0, c^)$ is locally structurally identifiable under these idealised conditions.*

Proof. We show that equality of the two output functions forces equality of each component in turn. Assume that

$$y_1(t) = y_2(t) \quad \text{for all observed } t.$$

Since $y(t) = \delta E(t)$ and δ is known, the exposed trajectory is determined directly by the output:

$$E(t) = \frac{y(t)}{\delta}.$$

Hence $E_1(t) = E_2(t)$.

Next, from

$$\frac{d\mathbf{l}}{dt} + \gamma \mathbf{l} = \delta \mathbf{E}(t),$$

and the known initial condition $\mathbf{l}(0) = \mathbf{l}_0$, the infectious trajectory is uniquely determined by $\mathbf{E}(t)$. Therefore

$$\mathbf{l}_1(t) = \mathbf{l}_2(t).$$

Define the infection flux

$$F(t) := \beta(t)(\rho \mathbf{E}(t) + \mathbf{l}(t)) \left(\frac{S(t)}{N} \right)^{1+\nu^2}.$$

From the exposed equation,

$$\frac{d\mathbf{E}}{dt} = F(t) - \delta \mathbf{E}(t),$$

so

$$F(t) = \frac{d\mathbf{E}}{dt} + \delta \mathbf{E}(t). \quad (2.6)$$

Thus $F(t)$ is also determined by the output $y(t)$. Since

$$\frac{dS}{dt} = -F(t),$$

and $S(0) = S_0$ is known, the susceptible trajectory is determined by

$$S(t) = S_0 - \int_0^t F(u) du.$$

Therefore identical outputs imply identical reconstructed trajectories

$$\mathbf{E}_1(t) = \mathbf{E}_2(t), \quad \mathbf{l}_1(t) = \mathbf{l}_2(t), \quad S_1(t) = S_2(t).$$

(i) *Early phase* ($t < t_0$): *identifiability of $\mathcal{R}_0$* . Before interventions, $c(t) = 1$, and at very early times $S(t) \approx N$, so $(S(t)/N)^{1+\nu^2} \approx 1$ for any ν . The $(\mathbf{E}, \mathbf{l})$ subsystem is then well approximated by the linear system

$$\frac{d}{dt} \begin{pmatrix} \mathbf{E} \\ \mathbf{l} \end{pmatrix} = A(\beta_0) \begin{pmatrix} \mathbf{E} \\ \mathbf{l} \end{pmatrix}, \quad \beta_0 = \frac{\mathcal{R}_0}{\rho/\delta + 1/\gamma},$$

where $A(\beta_0)$ depends on $\beta_0, \delta, \gamma, \rho$ but not on ν . Its dominant eigenvalue gives the early growth rate $r(\mathcal{R}_0)$ of $y(t) = \delta \mathbf{E}(t)$. For fixed (ρ, δ, γ) this map is monotone, hence invertible, so the early

growth rate determines $\mathcal{R}_0$ uniquely:

$$y_1(t) = y_2(t) \text{ for small } t \implies \mathcal{R}_{0,1} = \mathcal{R}_{0,2},$$

while ν and c^* are not informed by the very early epidemic.

(ii) *Late tail: identifiability of c^* .* Fix $(\mathcal{R}_0, \nu, t_0)$ common to both sets. We first make the tail behaviour of S precise. The susceptible equation is

$$\frac{dS}{dt} = -\beta c(t) \{\rho E(t) + I(t)\} \left(\frac{S(t)}{N} \right)^{1+\nu^2}. \quad (2.7)$$

Once $R_{\text{eff}}(t) < 1$, the exposed and infectious compartments decay towards zero through the linear transition and removal terms in the E - and I -equations in (2.1). The right-hand side of (2.7) is driven only by the residual incidence, which is proportional to $\rho E(t) + I(t)$; hence $dS/dt \rightarrow 0$ as $E(t), I(t) \rightarrow 0$, and $S(t)$ approaches a limiting value S_∞ . Over this late phase $S(t)$ therefore changes only through the small residual tail incidence, whereas $E(t)$ and $I(t)$ continue to decay through the exposed–infectious subsystem; this is the precise sense in which $S(t)$ is locally slowly varying in the tail. Given this, at fixed $(\mathcal{R}_0, \nu, t_0)$ a change in c^* rescales the effective transmission through $R_{\text{eff}}(t) \propto \mathcal{R}_0 c(t) (S(t)/N)^{\nu^2}$, altering the late-time incidence level and decay rate. Hence if $\mathcal{R}_{0,1} = \mathcal{R}_{0,2}$ and $\nu_1 = \nu_2$ but $c_1 \neq c_2$, the tails of y_1 and y_2 cannot coincide for all large t . Identical outputs therefore force

$$c_1 = c_2.$$

(iii) *Transition timing t_0 .* Assume $\mathcal{R}_{0,1} = \mathcal{R}_{0,2}$ and $c_1 = c_2 = c^*$. Then $c_i(t) = g(t - t_{0,i}; c^*)$ differ only by a horizontal shift of the fixed shape g . If $t_{0,1} \neq t_{0,2}$, the time at which $c_i(t)$ departs from 1 and approaches c^* differs, so the timing of the bending and peak of the incidence curve differs. Since g is not invariant under non-zero time shifts, identical outputs force

$$t_{0,1} = t_{0,2}.$$

(iv) *Full trajectory: identifiability of ν .* With $(\mathcal{R}_0, t_0, c^*)$ common, the only free parameter is ν . If $y_1(t) = y_2(t)$ for all t , then $E_1(t) = E_2(t)$ and, by the I -equation in (2.1), $I_1(t) = I_2(t)$. The two S -equations share the same $\beta(t)$ and the same (E, I) ; if the outputs coincide for all t then $S_1(t) = S_2(t)$, so

$$\left(\frac{S(t)}{N} \right)^{1+\nu_1^2} = \left(\frac{S(t)}{N} \right)^{1+\nu_2^2} \quad \text{for all } t.$$

Under a non-trivial epidemic $S(t)/N$ ranges over a continuum in $(0, 1]$. For fixed $x \in (0, 1)$ the map $\alpha \mapsto x^\alpha$ is injective, so equality over a continuum of x forces $1 + \nu_1^2 = 1 + \nu_2^2$, i.e. $\nu_1 = \nu_2$ (using $\nu > 0$).

Combining (i)–(iv), $\theta_1 = \theta_2$. □

Remark The argument is informal and relies on strong assumptions: continuous noise-free observation of $y(t)$ over the whole epidemic, known $(\rho, \delta, \gamma, N)$, a fixed NPI shape g , and known initial conditions. A rigorous proof using differential-algebra or differential-geometric tools (e.g. STRIKE-GOLDD [75] or `StructuralIdentifiability.jl` [77]) is beyond the present scope. The input–output argument suggests that the parameter vector is *structurally identifiable* under idealised noise-free observation; the remainder of the chapter shows that this does not guarantee *practical* identifiability from finite, noisy epidemic data.

2.3.2 Origin of the ν – c^* compensation ridge

The analytical basis of the ν – c^* compensation mechanism was derived in Section 1.8. In particular, equation (1.109) shows that, during the sustained-intervention phase of a single epidemic, the data may primarily constrain the composite transmission-reduction factor

$$c_{\text{eff}} \approx c^* \left(\frac{S^*}{N} \right)^{\nu^2}, \quad (2.8)$$

where S^*/N is a representative susceptible fraction over the relevant part of the epidemic trajectory.

Figure 1.7 in Chapter 1 shows combinations of (ν, c^*) that preserve a fixed value of c_{eff} . Here, Figure 2.3 presents the complementary view required for the identifiability analysis: it shows how c_{eff} changes with susceptible depletion for a fixed c^* and several values of ν .

For $\nu = 0$, the factor c_{eff} remains equal to c^* and does not depend on susceptible depletion. For $\nu > 0$, it decreases as S^*/N falls, with a stronger decline at larger values of ν . Thus, greater susceptibility heterogeneity can generate an additional reduction in transmission beyond that produced by the intervention alone.

Because $0 < S^*/N < 1$, increasing ν decreases $(S^*/N)^{\nu^2}$. An increase in ν can therefore be partly offset by an increase in c^* , corresponding to a weaker inferred intervention, while leaving c_{eff} approximately unchanged. This produces the positive ν – c^* association and the shallow likelihood ridge examined in the remainder of the chapter.

The approximation is local rather than exact because the susceptible fraction and the exposed

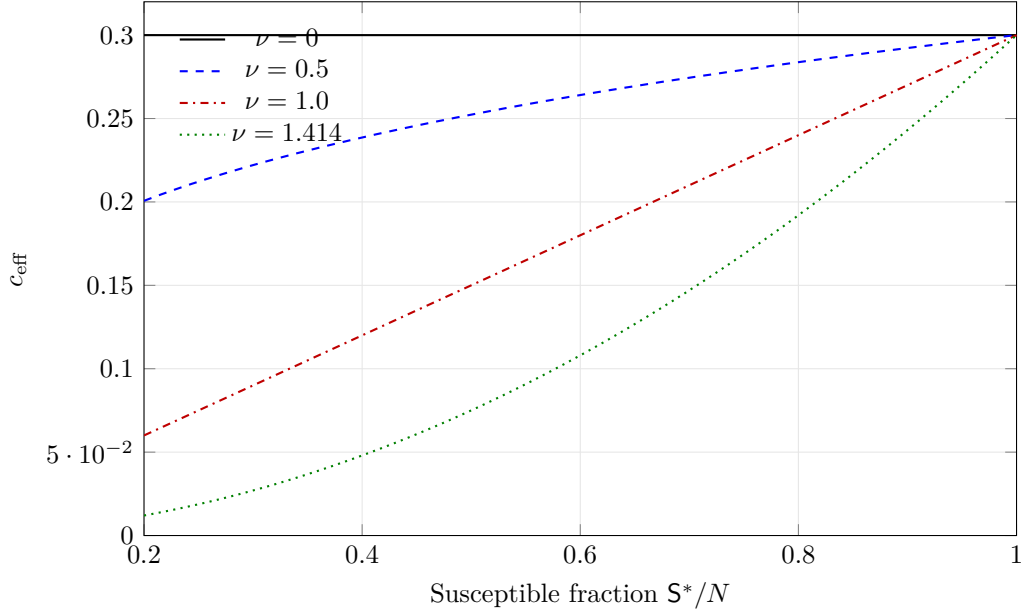


Figure 2.3: Effective transmission-reduction factor $c_{\text{eff}} \approx c^*(S^*/N)^{\nu^2}$ plotted as a function of the representative susceptible fraction S^*/N , for $c^* = 0.3$ and several values of ν . When $\nu = 0$, $c_{\text{eff}} = c^*$ is constant. For larger ν , the factor $(S^*/N)^{\nu^2}$ decreases more strongly as the susceptible fraction is depleted, so heterogeneity produces an additional reduction in effective transmission. Therefore, when fitting a single epidemic wave, an increase in ν can be compensated by an increase in c^* , corresponding to a weaker inferred intervention.

and infectious populations evolve throughout the epidemic. The following analyses therefore assess whether this compensation produces practical non-identifiability when $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$ is estimated from finite, noisy single-epidemic trajectories.

The next section introduces the four simulation scenarios used in the single-epidemic identifiability study.

2.4 Baseline Analysis: Practical Identifiability

This section investigates the main topic of the chapter: practical identifiability of the parameter vector $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$ using noisy data generated from the model equations (2.1). We present the simulation design as follows.

2.4.1 Simulations

We aim to identify the effects of individual variation and NPIs on epidemic trajectories based on a given epidemic dataset affected by both. We design a systematic study with four different incidence time series generated as follows:

S1: Homogeneous model ($\nu = 0$) without NPIs ($c^* = 1$).

S2: Heterogeneous model ($\nu > 0$) without NPIs ($c^* = 1$).

S3: Homogeneous model ($\nu = 0$) with NPIs ($c^* = 0.30$).

S4: Heterogeneous model ($\nu = 1.414$) with NPIs ($c^* = 0.30$).

All four deterministic baseline scenarios use the same initial epidemic seed, $I_0 = 40$ and $E_0 = 100$, so that differences between scenarios are driven by the presence or absence of susceptibility heterogeneity and NPIs, rather than by different starting epidemic sizes. The datasets are generated by running model (2.1), with NPIs as in (2.3). We define incidence of infection as $\delta E(t)$ and adopt a Poisson error structure to obtain the time series

$$y(t) \sim \text{Poisson}(\delta E(t)) \quad t = 1, \dots, T.$$

For each combination of parameters, 200 datasets¹ are simulated. We fit both the homogeneous model and the model with heterogeneity represented by parameter ν , using four fitted model specifications:

M1: Homogeneous, no NPIs: set $\nu = 0$ and $c(t) = 1$ for all t , and estimate $\mathcal{R}_0$ only;

M2: Heterogeneous, no NPIs: set $c(t) = 1$ for all t and estimate $(\mathcal{R}_0, \nu)$;

M3: Homogeneous, with NPIs: set $\nu = 0$, use the linear NPI profile (2.3), and estimate $(\mathcal{R}_0, t_0, c^*)$;

M4: Heterogeneous, with NPIs: use the linear NPI profile (2.3) and estimate $(\mathcal{R}_0, \nu, t_0, c^*)$.

All other epidemiological parameters are fixed at the values in Table 2.1. These labels are used throughout the baseline results (Table 2.2) to distinguish correctly specified fits, misspecified fits, and over-parameterised fits.

2.4.2 Typical epidemic trajectories under the four scenarios

Before presenting the estimation results, it is useful to visualise the four data-generating scenarios used in the baseline simulation study. Figure 2.4 shows representative simulated daily-incidence trajectories under homogeneous and heterogeneous susceptibility, with and without NPIs. The black curve in each panel is the deterministic incidence from the reduced SEIR model, while the coloured points show five Poisson realisations of the corresponding daily reports.

¹A sample size of 200 replicated datasets was chosen to balance Monte Carlo precision against computational cost. By the central limit theorem, the Monte Carlo standard error of a replicated-sample mean is $s/\sqrt{200}$, where s is the empirical standard deviation of the quantity being summarised, so the simulation error in reported means is about $0.071s$. For estimated coverage probabilities, the corresponding standard error is $\sqrt{\hat{p}(1-\hat{p})}/200$, which is at most 0.035 and is about 0.015 when $\hat{p} \approx 0.95$. Thus 200 replicates provide adequate precision for the distributional summaries reported here while remaining computationally feasible, since each replicate requires numerical ODE integration and iterative likelihood optimisation.

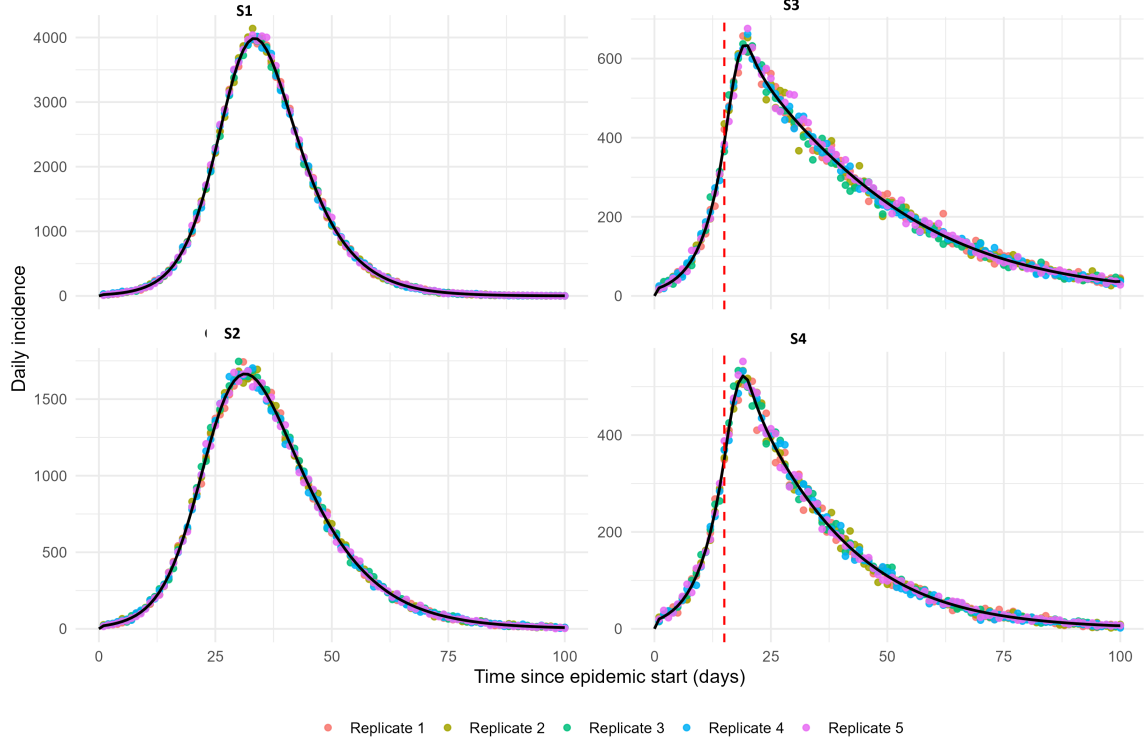


Figure 2.4: Typical epidemic trajectories under homogeneous and heterogeneous susceptibility, with and without a sustained piecewise-linear NPI profile. The black curve gives the deterministic model incidence, while coloured points show five simulated Poisson realisations of daily reports. The red dashed line marks the onset of contact reduction at $t_0 = 15$ days in the NPI scenarios. Heterogeneity lowers and broadens the epidemic peak through selective depletion, while NPIs reduce transmission through the contact multiplier $c(t)$. The vertical scales differ across panels to make the trajectory shapes visible.

2.5 Maximum likelihood estimation

With the heterogeneous susceptibility models, we estimate two sets of parameters from the datasets described in Section 2.4.1:

$$\theta_a := (\mathcal{R}_0, \nu) \quad (2.9)$$

$$\theta_b := (\mathcal{R}_0, \nu, t_0, c^*). \quad (2.10)$$

The first case corresponds to estimating θ_a by fitting the no-NPI models (**M1** and **M2**) to datasets **S1** and **S2**. This is to establish how well heterogeneity can be estimated in the absence of the potentially confounding effects of interventions, and to test how reliable the heterogeneous susceptibility model is at estimating “no variation” when the dataset has been generated by the homogeneous model.

The second case of interest is estimating θ_b by fitting the NPI models (**M3** and **M4**) to datasets **S3** and **S4**. We also fit **M3/M4** to **S1** and **S2** to test whether the inference correctly identifies the absence of interventions when $c^* = 1$.

Let $x_1, x_2, \dots, x_n$ be the model simulation (expected incidence) corresponding to the observed time series epidemic data $y_1, y_2, \dots, y_n$.

We assume that the observations y_i are conditionally independent given the model states x_i . The likelihood function is defined as the product of the individual probability mass functions:

$$\mathcal{L}(\boldsymbol{\theta}) := \prod_{i=1}^n f(y_i \mid x_i, \boldsymbol{\theta}),$$

where $f(\cdot)$ is the probability mass function (Poisson or Negative Binomial) and $\boldsymbol{\theta}$ is the vector of estimated parameters.

It is convenient to work with the log-likelihood function, given by:

$$\ell(\boldsymbol{\theta}) := \log \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(y_i \mid x_i, \boldsymbol{\theta}). \quad (2.11)$$

Parameter estimation was performed in RStudio by maximising (2.11), representing the probability of observing the actual data $\{y_i\}$ given the model simulations $\{x_i\}$. This was implemented by minimising the negative log-likelihood $-\ell(\boldsymbol{\theta})$ using the `optim()` function in R with the Nelder–Mead algorithm, with the maximum number of iterations set between 1 000 and 1 800 depending on the model complexity, and the default relative convergence tolerance of approximately 1.49×10^{-8} . All estimated parameters were transformed to an unconstrained scale before optimisation: a log-transformation for strictly positive parameters ($\mathcal{R}_0, \nu, t_0$) and a logit-transformation for the bounded parameter $c^* \in (0, 1)$. At convergence, estimates were back-transformed to the natural scale via the exponential and inverse-logit functions respectively. The Hessian matrix of the negative log-likelihood at the optimum was obtained numerically (finite differences) and, when positive definite, inverted to yield approximate 95% Wald confidence intervals on the transformed scale, which were then back-transformed.

In a small number of the 200 simulated datasets, the optimisation either did not converge or the Hessian was not positive definite. These cases were discarded from the summary statistics reported below.

2.6 Baseline analysis

This section presents a baseline simulation–inference study for a representative choice of parameters. We consider four simulated scenarios (**S1–S4**) described in Section 2.4.1, and fit the four models (**M1–M4**) described under Table 2.2. The aim is to assess (i) whether the fitted model

Table 2.2: Baseline simulation–inference combinations. Each row corresponds to a simulated scenario **S** x and a fitted model **M** y . Entries are representative MLEs with 95% Wald confidence intervals in brackets.

Simulated	Fitted	Parameter estimates (95% CI)				
		$\mathcal{R}_0$	ν	t_0	c^*	AIC
S1	M1	3.00 (3.00, 3.01)	NA	NA	NA	815.87
	M2	3.00 (2.99, 3.01)	0.067(0.017, 0.249)	NA	NA	816.90
	M3	3.00 (2.99, 3.02)	NA	0.001 [†]	0.999 (0.784, 1.00) [‡]	808.03
	M4	3.01 (2.99, 3.04)	$\approx 0^*$	0.11 [†]	0.997 (0.939, 0.999) [‡]	805.70
S2	M1	2.90 (2.89, 2.91)	NA	NA	NA	30237.71
	M2	3.00 (2.99, 3.01)	1.415 (1.40, 1.43)	NA	NA	811.76
	M3	2.74 (2.73, 2.75)	NA	28.48 (28.26, 28.69)	0.352 (0.348, 0.355)	2137.56
	M4	3.00 (2.98, 3.05)	1.411 (1.389, 1.423)	1.75 (1.02, 2.30) [†]	0.983 (0.893, 1.00) [‡]	824.10
S3	M3	3.00 (2.97, 3.04)	NA	15.0 (14.4, 15.60)	0.30 (0.296, 0.304)	791.09
	M4	3.00 (2.97, 3.04)	0.201 (0.04, 0.912)	14.92 (14.32, 15.56)	0.302 (0.296, 0.312)	793.66
S4	M3	2.94 (2.89, 3.00)	NA	14.91 (14.3, 15.5)	0.24 (0.236, 0.243)	728.57
	M4	3.00 (2.95, 3.05)	1.406 (1.157, 1.768)	15.00 (14.37, 15.65)	0.299 (0.274, 0.331)	712.87

Simulation scenarios:

S1: homogeneous ($\nu = 0$), no NPIs ($c^* = 1$); **S2:** heterogeneous ($\nu = 1.414$), no NPIs ($c^* = 1$); **S3:** homogeneous ($\nu = 0$), with NPIs ($c^* = 0.3$); **S4:** heterogeneous ($\nu = 1.414$), with NPIs ($c^* = 0.3$).

Fitted models:

M1: homogeneous, no NPIs (fit $\mathcal{R}_0$); **M2:** heterogeneous, no NPIs (fit $\mathcal{R}_0, \nu$); **M3:** homogeneous, with NPIs (fit $\mathcal{R}_0, t_0, c$); **M4:** heterogeneous, with NPIs (fit $\mathcal{R}_0, \nu, t_0, c$).

[†]When the simulated data have no intervention ($c^* = 1$), the change-point time t_0 is redundant because the transmission modifier does not change over time. In these cases the log-likelihood surface is effectively flat in the t_0 direction, so inverse-Hessian (Wald) confidence intervals for t_0 can become redundant or unstable.

[‡]When c^* is at or near the boundary ($c^* \approx 1$), symmetric Wald intervals on the natural scale are known to be poorly calibrated. We therefore interpret these fits primarily through whether the optimizer places c^* close to 1 (no intervention) and summarize uncertainty calibration across replicated datasets in Section 2.6.5.

Row (S1,M3) and (S1,M4):

For row (S1,M3) and (S1,M4), we have *Median $\hat{\nu} = 5.93 \times 10^{-6}$; the 95% CI spans several hundred orders of magnitude and is omitted as uninformative. Only 40 of 200 datasets (M4) and 70 of 200 (M3) yielded a positive-definite Hessian; summaries are computed over these valid fits.

class is correctly selected by relative fit quality (AIC) and (ii) how parameter estimates behave when the fitted model is correctly specified, misspecified, or over-parameterised.

For convenience, we also group simulated scenarios into two broader settings:

Case I: Data simulated under homogeneous susceptibility ($\nu = 0$) with $\mathcal{R}_0 = 3$, either without interventions ($c^* = 1$, scenario **S1**) or with interventions ($c^* = 0.3$, scenario **S3**).

Case II: Data simulated under heterogeneous susceptibility ($\nu = 1.414$) with $\mathcal{R}_0 = 3$, either without interventions ($c^* = 1$, scenario **S2**) or with interventions ($c^* = 0.3$, scenario **S4**).

Table 2.2 reports MLE summaries and 95% Wald confidence intervals. In scenarios with $c^* = 1$ (no intervention), t_0 is redundant and its Wald interval can become numerically unstable (marked [†]). When c^* is near the boundary $c^* \approx 1$, Wald intervals on the natural scale can also be poorly behaved and should be interpreted cautiously (marked [‡]).

2.6.1 Homogeneous datasets without NPIs (S1)

We first fit the no-NPI models to assess whether heterogeneity is correctly inferred as absent when the data are truly homogeneous.

Fitting **M1** (homogeneous, no NPIs) and **M2** (heterogeneous, no NPIs) to **S1** yields essentially identical estimates of $\mathcal{R}_0$ and very similar AIC values (rows **(S1,M1)** and **(S1,M2)** in Table 2.2). The heterogeneous model returns $\hat{\nu}$ close to 0 and the AIC difference between **M1** and **M2** is small, indicating no strong evidence that the added heterogeneity parameter improves fit for truly homogeneous data.

We next fit the NPI models **M3** and **M4** to the same **S1** datasets (rows **(S1,M3)** and **(S1,M4)**). In this over-parameterised setting (since the true data have $c^* = 1$), the models place $\hat{c}^*$ close to 1, as expected, but the change-point time $\hat{t}_0$ becomes redundant and its Wald interval becomes meaningless (marked $\dagger$). The AIC values remain close across **M3** and **M4** in this scenario, so the fitted results support the conclusion that the data do not require an intervention component.

2.6.2 Homogeneous datasets with NPIs (S3)

This scenario is informative because both heterogeneity and NPIs can flatten epidemic curves.

Fitting the NPI models **M3** (homogeneous with NPIs) and **M4** (heterogeneous with NPIs) to **S3** yields very similar estimates for $\mathcal{R}_0$, t_0 , and c^* (rows **(S3,M3)** and **(S3,M4)** in Table 2.2) which on average matches the specified values in Table 2.1. However, although the true data are homogeneous ($\nu = 0$), the heterogeneous model **M4** attributes part of the epidemic slowdown to heterogeneity, returning a non-zero $\hat{\nu}$ with a wide confidence interval. However the AIC values for **(S3,M3)** and **(S3,M4)** are also close, indicating goodness-of-fit alone may not reliably distinguish whether flattening is due to NPIs or heterogeneity in this single-epidemic setting. Thus, although the heterogeneous model returns a non-zero estimate of ν , this does not lead to substantial bias in the estimates of t_0 or c^* .

The fitted trajectories for **S3** are shown in Figure 2.6(a). Both the homogeneous and heterogeneous NPI-augmented models reproduce the observed epidemic trajectory closely. However, because the data are generated without susceptibility heterogeneity, any non-zero heterogeneity inferred by the heterogeneous model should be interpreted as redundant flexibility rather than evidence of true population variation. This scenario therefore provides a useful check on whether the model can avoid attributing an intervention-generated slowdown to heterogeneity.

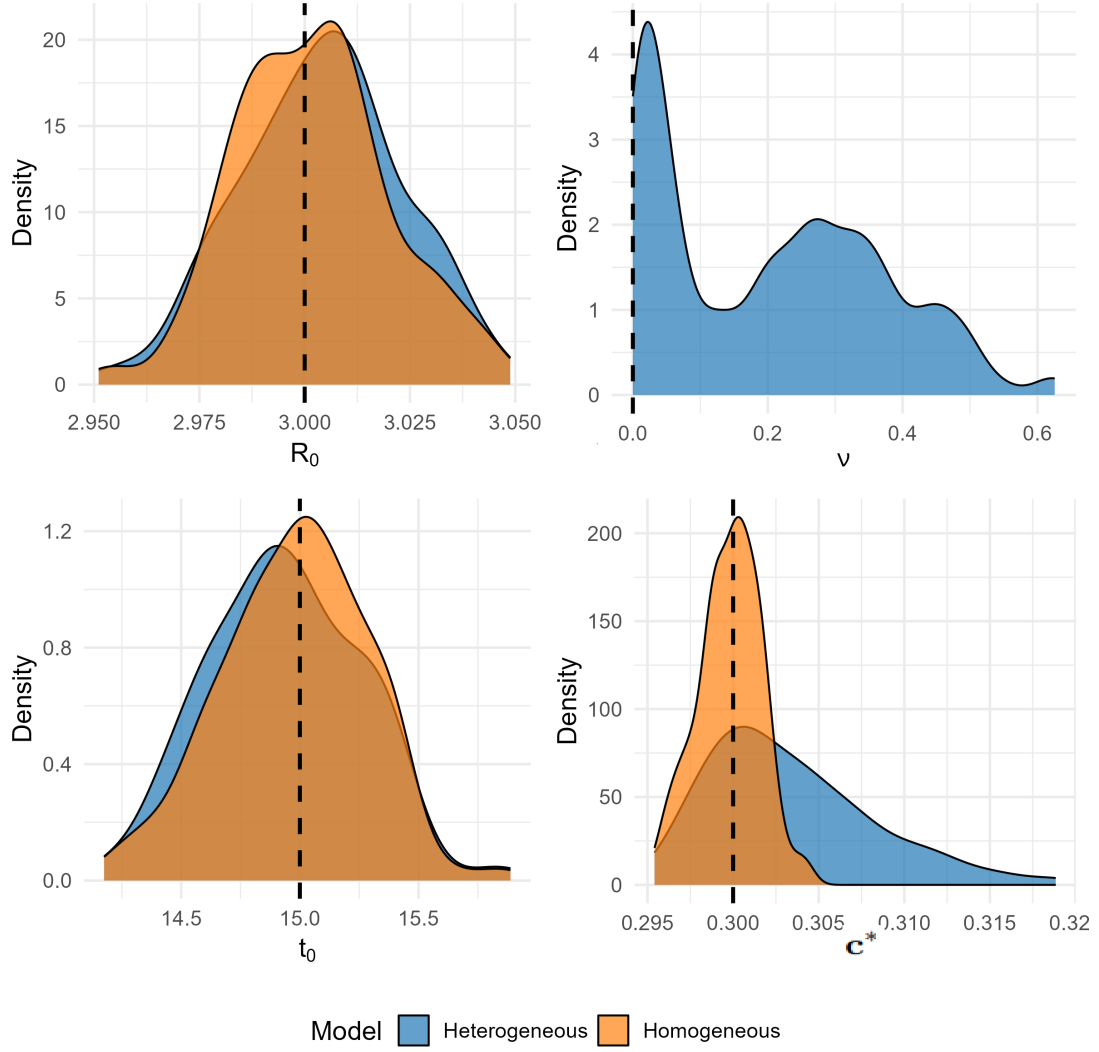
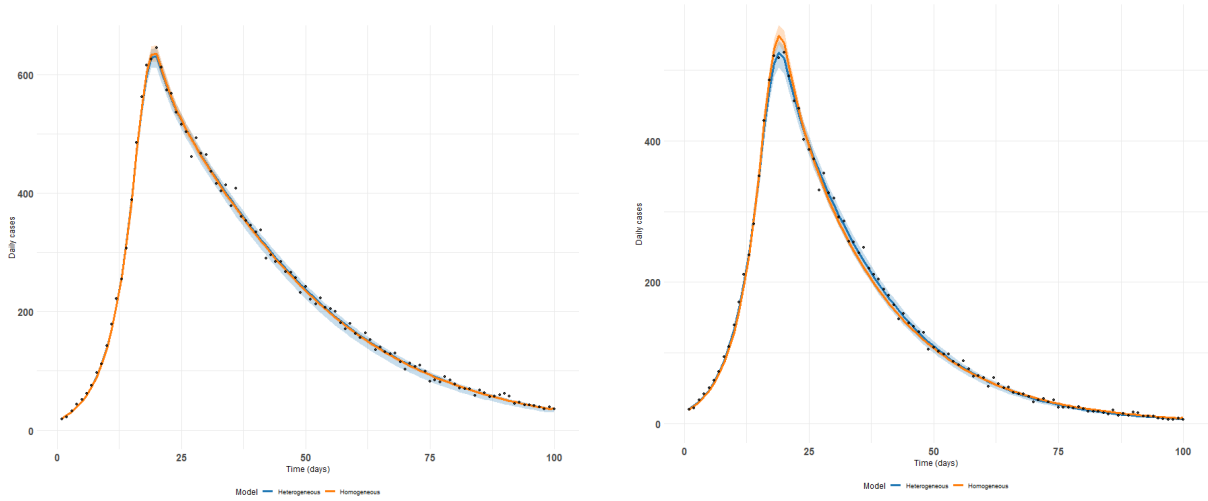


Figure 2.5: Marginal distributions of maximum-likelihood estimates across replicated fits for Scenario **S3**, where data are generated from the homogeneous model with NPIs. The data-generating values are $\mathcal{R}_0 = 3$, $t_0 = 15$ days, $c^* = 0.3$, and $\nu = 0$. Blue represents estimates from the heterogeneous susceptibility model **M4**, while orange represents estimates from the homogeneous model **M3**. Vertical dashed lines indicate the data-generating values. Both models recover $\mathcal{R}_0$, t_0 , and c^* accurately, but the heterogeneous model produces a right-skewed distribution for $\hat{\nu}$. Since the true data are homogeneous, this non-zero $\hat{\nu}$ should be interpreted as redundant flexibility rather than evidence of true susceptibility heterogeneity.

2.6.3 Heterogeneous datasets without NPIs (S2)

We first fit the no-NPI models to assess whether heterogeneity is recoverable when interventions are absent.

Comparing rows **(S2,M1)** and **(S2,M2)** in Table 2.2, the heterogeneous no-NPI model **M2** accurately recovers both $\mathcal{R}_0$ and ν and achieves a much lower AIC than the homogeneous no-NPI model **M1**. The homogeneous model attempts to compensate for the missing heterogeneity by shifting $\mathcal{R}_0$ downward from the specified value of 3, but the resulting fit remains poor, as



(a) Scenario **S3**: homogeneous data with NPIs

(b) Scenario **S4**: heterogeneous data with NPIs

Figure 2.6: Model fits for the two NPI scenarios. Panel (a) shows Scenario **S3**, where data are generated from the homogeneous model with NPIs. Panel (b) shows Scenario **S4**, where data are generated from the heterogeneous model with NPIs. In both panels, the homogeneous and heterogeneous NPI-augmented models reproduce the observed epidemic trajectory closely. The key difference is interpretive: in **S3**, any inferred heterogeneity is redundant because the data-generating process is homogeneous, whereas in **S4** both heterogeneity and intervention effects are genuinely present and can compensate for one another in the likelihood.

reflected by the very large AIC for (**S2,M1**).

We next fit the NPI models (**M3** and **M4**) to the same **S2** datasets (rows (**S2,M3**) and (**S2,M4**)). Model **M4** still recovers $\mathcal{R}_0$ and ν well and estimates c^* close to 1, indicating no intervention, while the inferred t_0 is redundant (marked †). In contrast, the homogeneous NPI model **M3** introduces an apparent intervention (estimating $c^* < 1$ and a finite t_0), thereby mis-attributing heterogeneity-driven flattening to NPIs. Nevertheless, the AIC remains substantially worse for (**S2,M3**) than for (**S2,M4**), indicating that allowing heterogeneity is strongly favoured when the true data are heterogeneous even if the fitted model also permits NPIs.

2.6.4 Heterogeneous datasets with NPIs (**S4**)

This is the full model setting (heterogeneous susceptibility with interventions). The fitted trajectories for Scenario **S4** are shown in Figure 2.6(b). As in Scenario **S3**, both fitted models can reproduce the observed incidence trajectory. However, in this case the similarity of the fitted curves is more consequential, because the data are generated by two mechanisms that act in the same direction. The visual overlap therefore motivates the profile-likelihood analysis below, which examines how changes in ν can be compensated by changes in c^* while preserving a similar epidemic curve. Comparing rows (**S4,M3**) and (**S4,M4**) in Table 2.2, the heterogeneous NPI model **M4** recovers all parameters simultaneously with estimates close to the specified simulation

values, and achieves the lower AIC. The homogeneous NPI model **M3** underestimates $\mathcal{R}_0$ and compensates by estimating a stronger intervention (smaller c^*), illustrating the same trade-off mechanism seen in **S2** under misspecification.

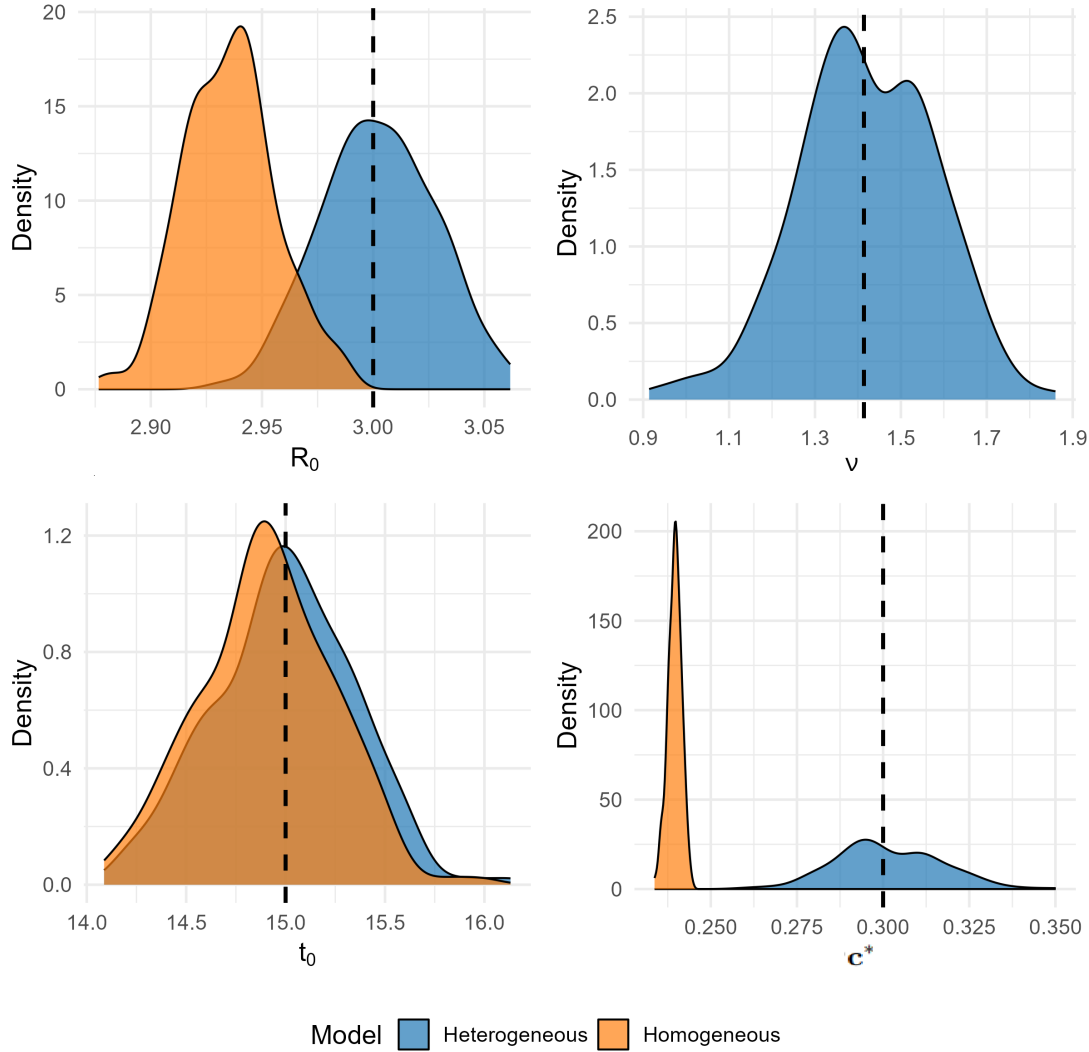


Figure 2.7: Marginal distributions of maximum-likelihood estimates across replicated fits for Scenario **S4**, where data are generated from the heterogeneous-susceptibility model with NPIs. The data-generating values are $\mathcal{R}_0 = 3$, $\nu = 1.414$, $t_0 = 15$ days, and $c^* = 0.3$. Blue represents estimates from the heterogeneous susceptibility model **M4**, while orange represents estimates from the homogeneous model **M3**. Vertical dashed lines indicate the data-generating values. The heterogeneous model recovers the four fitted parameters close to their data-generating values. By contrast, the homogeneous model underestimates $\mathcal{R}_0$ and estimates a smaller c^* , corresponding to a stronger inferred intervention, because it cannot represent selective depletion through susceptibility heterogeneity.

2.6.5 Uncertainty quantification in S3 and S4 across replicated datasets

The marginal distribution in Figure 2.7 summarises uncertainty in each fitted parameter separately. However, the practical-identifiability problem in Scenario **S4** is inherently joint: an estimation

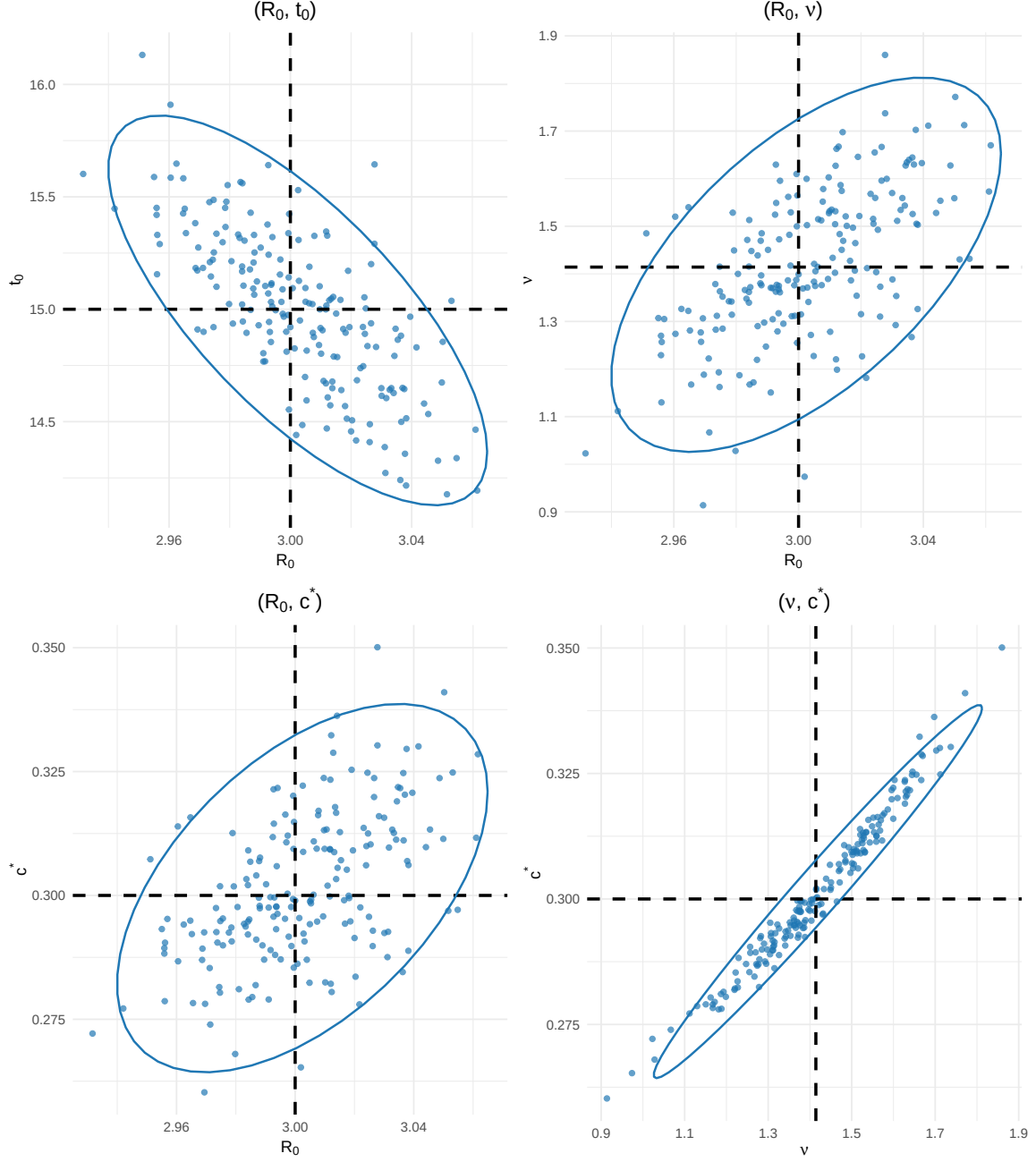


Figure 2.8: Empirical joint distributions of selected maximum-likelihood estimates across replicated single-epidemic fits under Scenario **S4**, where both heterogeneous susceptibility and NPIs are present. Each point corresponds to one simulated dataset, and the ellipses summarise the empirical point clouds. Dashed vertical and horizontal lines mark the data-generating values. The panels show $(\hat{\mathcal{R}}_0, \hat{t}_0)$, $(\hat{\mathcal{R}}_0, \hat{v})$, $(\hat{\mathcal{R}}_0, \hat{c}^*)$, and $(\hat{v}, \hat{c}^*)$. The strongest visual pattern is the positive association between $\hat{v}$ and $\hat{c}^*$, showing the compensation direction: larger inferred susceptibility heterogeneity can be accompanied by larger c^* , corresponding to weaker inferred contact reduction, while producing similar epidemic trajectories.

error in one parameter can be associated with a compensating error in another. To display this dependence directly, Figure 2.8 shows empirical joint distributions of selected maximum-likelihood estimates across the replicated **S4** datasets.

The strongest pattern is seen in the $(\hat{v}, \hat{c}^*)$ panel. Larger estimates of susceptibility heterogene-

Table 2.3: Median parameter correlations for the NPI scenarios, derived from the inverse Hessian matrix across replicated fits. Scenario **S3** corresponds to homogeneous data with NPIs, while Scenario **S4** corresponds to heterogeneous data with NPIs. The heterogeneous model estimates $(\mathcal{R}_0, \nu, t_0, c^*)$, whereas the homogeneous model estimates $(\mathcal{R}_0, t_0, c^*)$. Diagonal correlations are omitted because they are equal to one.

Parameter pair	Scenario S3 :		Scenario S4 :	
	M4	M3	M4	M3
$\mathcal{R}_0-t_0$	-0.949	-0.947	-0.726	-0.938
$\mathcal{R}_0-\nu$	0.129	—	0.594	—
$\mathcal{R}_0-c^*$	-0.192	-0.858	0.523	-0.656
$t_0-\nu$	-0.156	—	0.006	—
t_0-c^*	0.119	0.720	0.057	0.437
$\nu-c^*$	0.885	—	0.983	—

ity are associated with larger estimates of c^* , corresponding to weaker inferred contact reduction. This is the compensation direction predicted by the local approximation in Section 2.3.1: stronger selective depletion can be offset by a weaker sustained intervention while preserving a similar incidence trajectory.

The same pattern is supported by the inverse-Hessian correlations in Table 2.3. For Scenario S4 under the heterogeneous fitted model M4, the $\nu-c^*$ correlation is 0.983, which is the largest absolute correlation among the reported parameter pairs. The next largest correlations are the $\mathcal{R}_0-t_0$ correlation (-0.726), the $\mathcal{R}_0-\nu$ correlation (0.594), and the $\mathcal{R}_0-c^*$ correlation (0.523). Thus the dominant local compensation direction in the full heterogeneous-with-NPI setting is the $\nu-c^*$ direction. Together with the empirical joint distributions in Figure 2.8, however, they show that the main single-epidemic uncertainty is not only the marginal width of the estimate of ν or c^* , but the fact that these two parameters move together along a compensation ridge.

Furthermore, for each dataset we compute within-dataset Wald uncertainty using the inverse observed information matrix, obtained by inverting the Hessian of the negative log-likelihood evaluated at the MLE (in a transformed parameterization to respect parameter constraints) and reporting two diagnostics aggregated across replicated datasets: (i) the relative bias of the median MLE across replicates, and (ii) a typical within-dataset 95% Wald interval width (mean UCL-LCL across replicates).

Because the main inferential difficulty studied here concerns confounding between NPIs and heterogeneity, we report these diagnostics only for the two baseline scenarios with NPIs: S3 and S4; see Table 2.4.

Table 2.4: Uncertainty and bias summaries across 200 replicated datasets for the two NPI baseline scenarios: **S3** (homogeneous model with NPIs) and **S4** (heterogeneous model with NPIs). For each fitted model, SD is the standard deviation of the MLE across datasets (between-dataset variability). CI width is the mean 95% Wald interval width (UCL–LCL) computed within each dataset and averaged across datasets (within-dataset uncertainty). Bias is computed from the median MLE across datasets; relative bias is reported as a percentage of the true specified value where defined, otherwise absolute bias is reported.

Simulated	True model	Parameter	M3 (HOM)			M4 (HET)		
			Bias	SD	CI width	Bias	SD	CI width
S3	Hom. + NPIs	$\mathcal{R}_0$	0.491%	0.018	0.069	0.492%	0.018	0.073
		ν	–	–	–	0.263 [†]	0.154	0.622
		t_0	1.629%	0.307	1.141	1.685%	0.304	1.215
		c^*	0.485%	0.002	0.007	1.767%	0.005	0.017
S4	Het. + NPIs	$\mathcal{R}_0$	2.145%	0.021	0.073	0.687%	0.025	0.097
		ν	–	–	–	9.071%	0.160	0.605
		t_0	1.847%	0.336	1.166	1.860%	0.352	1.287
		c^*	20.160%	0.002	0.008	4.051%	0.015	0.057

[†]For **S3**, the true $\nu = 0$, so relative bias is undefined. We therefore report the absolute bias $|\text{median}(\hat{\nu}) - 0| = 0.263$ instead of a percentage.

2.6.6 Visualising practical non-identifiability along the ν – c^* ridge

The heterogeneous-with-NPI scenario is the most informative case for diagnosing the main single-epidemic identifiability problem, because both selective depletion and contact reduction are active. To visualise the resulting compensation geometry, we profiled the likelihood in ν for one representative simulated dataset from Scenario **S4**, while tracking the fitted value of c^* . The 95% profile-likelihood ridge was defined using the likelihood-ratio criterion

$$2\{\ell(\hat{\boldsymbol{\theta}}) - \ell_{\text{prof}}(\nu)\} \leq \chi_{1,0.95}^2.$$

Five representative points were then selected along this ridge and the corresponding daily-incidence trajectories were reconstructed.

Figure 2.9 shows that the fitted trajectories are almost indistinguishable over the observed epidemic window, even though the selected ridge points span a widerange of heterogeneity values. In this representative dataset, the selected values range from $\nu = 1.13$ to $\nu = 1.63$, with the corresponding fitted NPI floor increasing from $c^* = 0.27$ to $c^* = 0.32$. Thus, larger heterogeneity in susceptibility is accompanied by a weaker inferred intervention, while preserving almost the same incidence curve. This is the practical expression of the ν – c^* compensation mechanism derived above.

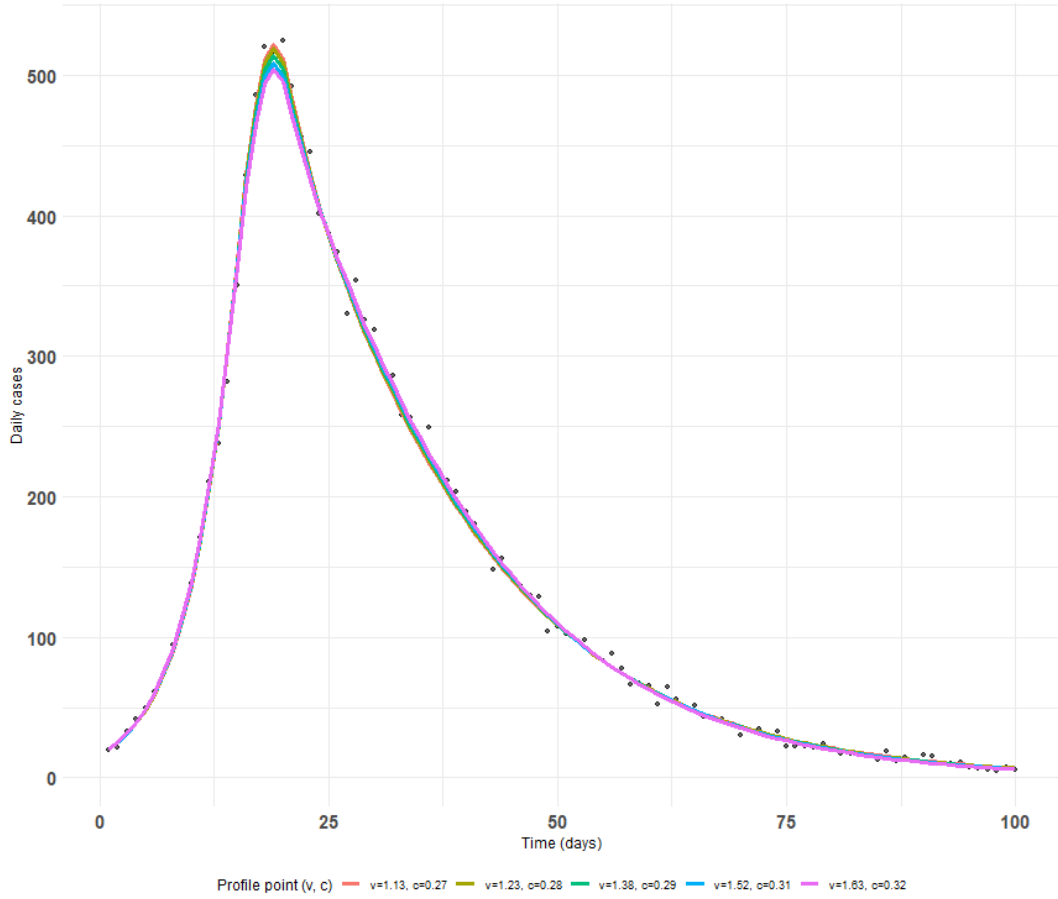


Figure 2.9: Fitted daily-incidence trajectories at five representative points along the 95% profile-likelihood ridge for ν , tracking c^* , in the heterogeneous-with-NPI scenario. Grey points show the mean observed daily reports from three simulated replicate datasets. The selected ridge points range from $\nu = 1.13$, $c^* = 0.27$ to $\nu = 1.63$, $c^* = 0.32$. Despite this variation in ν and c^* , the reconstructed incidence curves are visually very similar, illustrating that a single epidemic time series mainly constrains a composite reduction in transmission rather than uniquely separating heterogeneity in susceptibility from contact reduction.

2.7 Robustness to stochastic process noise

The baseline analysis in this chapter uses deterministic ordinary differential equation (ODE) solutions with Poisson observation error to assess practical identifiability of heterogeneity ν and NPI strength c^* . In finite populations, epidemic transmission is also affected by demographic stochasticity, or process noise, because infection, progression and recovery events occur as discrete random events rather than as deterministic flows. Here, we extend the baseline analysis of section 2.4, by implementing tau-leaping simulations, a discrete-state stochastic algorithm that explicitly accounts for demographic stochasticity (process noise). We repeated the simulation with heterogeneous susceptibility with NPIs; scenario S4, using an event-based stochastic simulator based on the τ -leap approximation as described in Chapter 8 of Bjørnstad [81].

The stochastic analysis therefore asks whether the same practical identifiability difficulty

remains when the data are generated from a discrete-state epidemic process.

Following the τ -leap approximation described by Bjørnstad [81], over a sufficiently short time step Δt , the number of events is approximated by Poisson random variables with means given by the event rates multiplied by Δt . This approximates exact event-by-event simulation while remaining computationally faster. Unlike the deterministic ODE solution, the τ -leaping trajectories fluctuate around the deterministic limit because the realised number of events in each time step is random.

Figure 2.10 illustrates this difference for the heterogeneous-with-NPI model at $N = 10,000$. The deterministic trajectory provides the smooth limiting behaviour, while the stochastic realisations show the additional variation introduced by process noise.

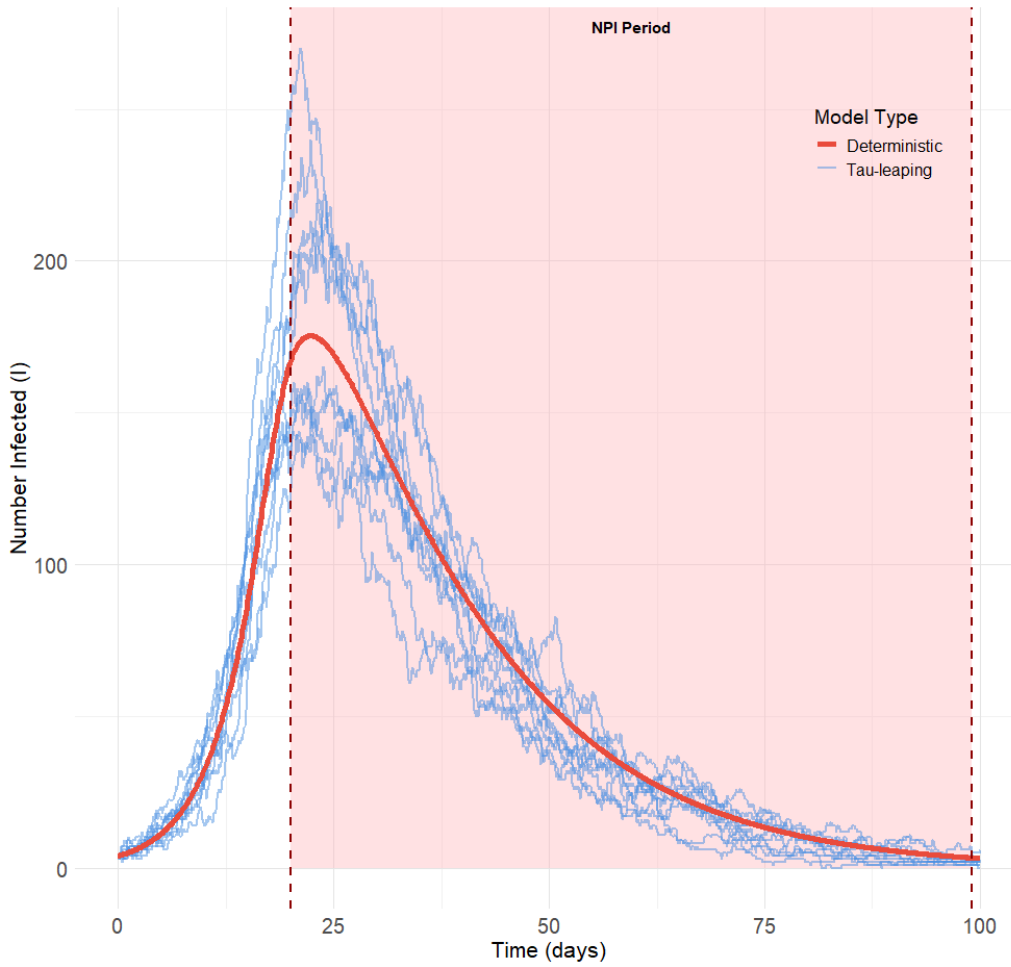


Figure 2.10: Comparison of deterministic and stochastic epidemic trajectories under the heterogeneous-with-NPI scenario at $N = 10,000$. The red curve represents the deterministic ODE limit. The blue trajectories are stochastic τ -leaping realisations. Their spread illustrates the variability introduced by process noise. The shaded region indicates the period over which the NPI profile is active.

2.7.1 Event-based stochastic simulation

Let

$$X(t) = (S(t), E(t), I(t), R(t), C(t))$$

denote the stochastic epidemic state, where $C(t)$ is a cumulative-incidence counter. The event hazards are

$$\lambda_{SE}(t) = \beta(t) (\rho E(t) + I(t)) \left(\frac{S(t)}{N} \right)^{1+\nu^2}, \quad (2.12)$$

$$\lambda_{EI}(t) = \delta E(t), \quad (2.13)$$

$$\lambda_{IR}(t) = \gamma I(t). \quad (2.14)$$

These correspond respectively to infection $S \rightarrow E$, progression $E \rightarrow I$, and removal $I \rightarrow R$. The progression event also increments the cumulative-incidence counter $C(t)$, so that daily incidence is obtained from increments in $C(t)$.

The event matrix has rows corresponding to events and columns corresponding to S, E, I, R, C :

$$M = \begin{pmatrix} -1 & +1 & 0 & 0 & 0 \\ 0 & -1 & +1 & 0 & +1 \\ 0 & 0 & -1 & +1 & 0 \end{pmatrix}.$$

The time-varying transmission rate is computed from the same NPI multiplier used in the deterministic analysis,

$$\beta(t) = \frac{\mathcal{R}_0 c(t)}{\rho/\delta + 1/\gamma}. \quad (2.15)$$

For a small step size Δt , the hazards are treated as approximately constant over $(t, t + \Delta t]$.

The update proceeds as follows:

1. Evaluate the hazards $\lambda_j(t)$ and the total rate $r(t) = \sum_j \lambda_j(t)$.
2. Draw the total number of events

$$K \sim \text{Poisson}\{r(t)\Delta t\}.$$

3. Conditional on K , allocate event types using probabilities $\lambda_j(t)/r(t)$, equivalently by a multinomial draw.
4. Update the state by adding the corresponding rows of the event matrix M , with non-

negativity enforced where necessary.

Synthetic daily incidence is computed as

$$\Delta C(t) = C(t) - C(t-1).$$

The simulated stochastic incidence trajectories are then fitted using the same maximum-likelihood framework as in Section 2.5. We fit both the homogeneous NPI model **M3**, estimating $(\mathcal{R}_0, t_0, c^*)$, and the heterogeneous NPI model **M4**, estimating $(\mathcal{R}_0, \nu, t_0, c^*)$. Approximate Wald confidence intervals are obtained from the inverse Hessian when the fitted Hessian is positive definite.

2.7.2 Simulation design

We simulated 200 independent stochastic trajectories for each of three population sizes,

$$N \in \{10,000, 100,000, 5,000,000\}.$$

These represent small, medium and large epidemic populations. The data-generating parameters were those of Scenario **S4**: Table 2.1.

Initial conditions were scaled proportionally with population size using $l_0 = 0.0004N$, $E_0 = 2.5l_0$ to maintain the initial infected and exposed fractions comparable across the three population sizes.

2.7.3 Results

Table 2.5 summarises the median parameter estimates and 95% Wald confidence intervals across the 200 stochastic replicates for each population size and fitted model. Figure 2.11 shows a representative fitted trajectory from the stochastic analysis for the baseline analysis of Scenario **S4**. Both fitted models track the observed epidemic curve, but they attribute the slowdown differently. The heterogeneous model can use selective depletion through ν , whereas the homogeneous model has to absorb the missing heterogeneity mainly through the fitted contact multiplier c^* .

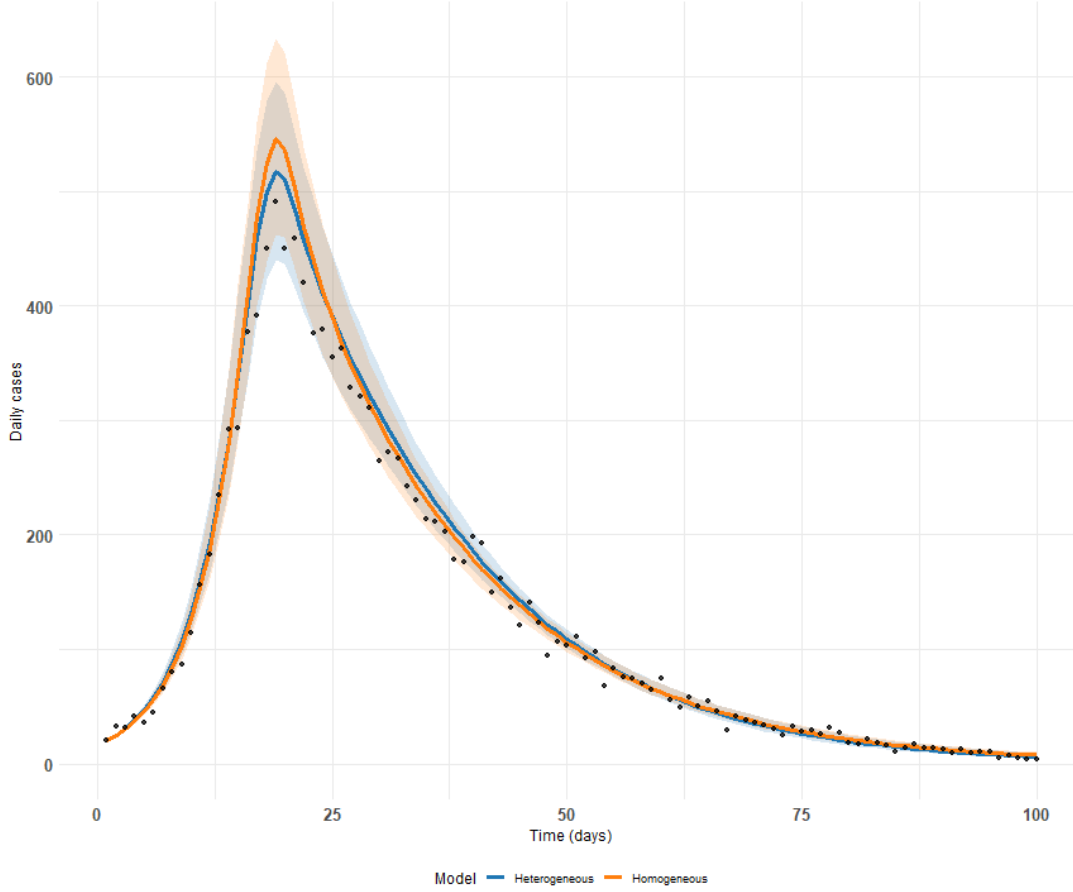


Figure 2.11: Representative fit to a stochastic τ -leaping realisation under Scenario **S4**. Black points show the simulated daily incidence counts. The blue curve and band show the heterogeneous fit **M4**, while the orange curve and band show the homogeneous fit **M3**. Process noise perturbs the observed trajectory around the deterministic epidemic shape, while preserving the same model- comparison problem: homogeneous and heterogeneous fits can both track the observed curve but attribute the epidemic slowdown differently.

Table 2.5: Parameter estimates from τ -leaping simulations of Scenario **S4** across three population sizes. Entries are medians with 95% Wald confidence intervals in parentheses across 200 stochastic replicates. The data-generating values are $\mathcal{R}_0 = 3$, $\nu = 1.414$, $t_0 = 15$, and $c^* = 0.3$. **M3**: homogeneous model with NPIs; **M4**: heterogeneous model with NPIs.

Population	Fitted model	$\mathcal{R}_0$	ν	t_0	c^*	AIC
10,000	M3	2.87 (2.75, 3.00)	–	15.30 (13.56, 17.28)	0.247 (0.235, 0.261)	474.80
	M4	2.95 (2.82, 3.10)	1.65 (1.20, 2.46)	15.52 (13.62, 17.91)	0.339 (0.269, 0.416)	469.80
100,000	M3	2.91 (2.87, 2.94)	–	15.20 (14.60, 15.80)	0.243 (0.239, 0.247)	732.90
	M4	2.98 (2.94, 3.03)	1.49 (1.22, 1.82)	15.27 (14.64, 15.92)	0.309 (0.281, 0.338)	711.40
5,000,000	M3	2.93 (2.93, 2.94)	–	14.92 (14.84, 15.00)	0.240 (0.239, 0.240)	1936.00
	M4	3.00 (2.99, 3.01)	1.42 (1.37, 1.46)	15.00 (14.91, 15.09)	0.300 (0.296, 0.305)	1106.00

The deterministic baseline analysis in Figure 2.8 showed that the replicated estimates are not independent across parameters. In particular, the $(\hat{\nu}, \hat{c}^*)$ panel displayed a strong positive association, indicating that larger inferred susceptibility heterogeneity can be offset by weaker inferred sustained contact reduction. The same diagnostic can be applied to the

stochastic τ -leaping fits. This provides a direct comparison between the deterministic baseline and the process-noise setting: if the same pairwise structure appears under τ -leaping, then the compensation geometry is not an artefact of using smooth ODE-generated trajectories.

Figure 2.12 shows this stochastic analogue for the $N = 100,000$ τ -leaping experiment. Compared with the deterministic baseline, the point clouds are broader because each replicate contains demographic process noise. However, the main dependence pattern is preserved. The $(\hat{\nu}, \hat{c}^*)$ panel again shows a clear positive diagonal direction, confirming that the practical identifiability problem remains present when the epidemic data are generated from a discrete stochastic process.

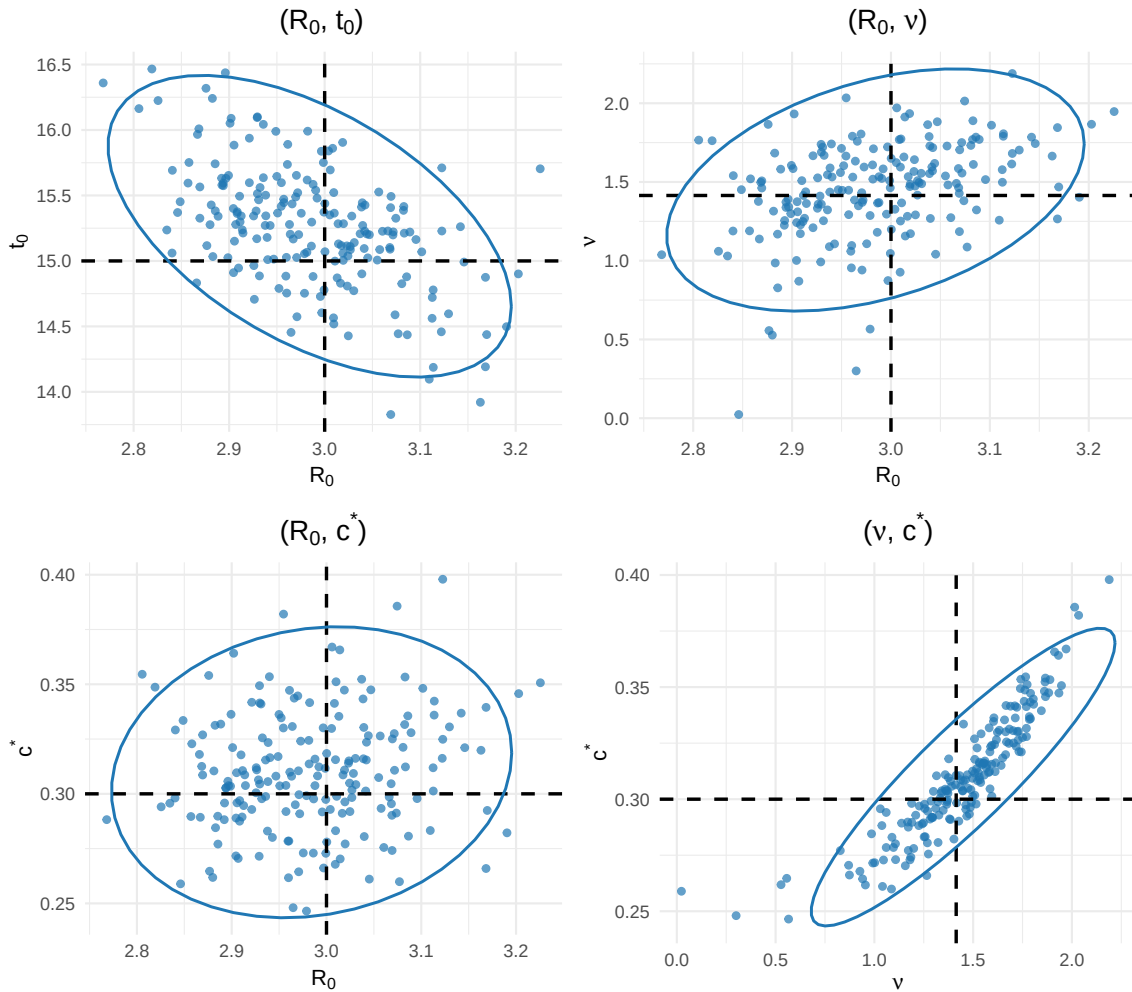


Figure 2.12: Empirical joint distributions of selected maximum-likelihood estimates across replicated stochastic τ -leaping fits under Scenario **S4**, shown here for $N = 100,000$. Each point corresponds to one stochastic epidemic replicate fitted with the heterogeneous model **M4**. Dashed vertical and horizontal lines mark the data-generating values $(\mathcal{R}_0, \nu, t_0, c^*) = (3, 1.414, 15, 0.3)$. The ellipses provide visual summaries of the empirical point clouds and should not be interpreted as likelihood contours. The panels show $(\hat{\mathcal{R}}_0, \hat{t}_0)$, $(\hat{\mathcal{R}}_0, \hat{\nu})$, $(\hat{\mathcal{R}}_0, \hat{c}^*)$, and $(\hat{\nu}, \hat{c}^*)$. The lower-right panel reproduces the main compensation direction seen in the deterministic baseline analysis: larger inferred susceptibility heterogeneity is associated with larger c^* , corresponding to weaker inferred sustained contact reduction. Thus, demographic process noise broadens the replicated estimates but does not remove the ν - c^* ridge.

The stochastic joint distributions therefore support the same interpretation as the deterministic baseline. Process noise increases the spread of the replicated estimates, especially in the $\mathcal{R}_0$, t_0 , and ν directions, but the orientation of the main ridge is unchanged. This is important because it shows that the single-epidemic difficulty is not created by the deterministic simulation method. Rather, both deterministic and stochastic data-generating mechanisms allow a family of parameter combinations in which susceptibility heterogeneity and sustained contact reduction compensate for one another.

2.7.4 Precision gains with increasing population size

Table 2.5 shows that uncertainty narrows systematically as population size increases. To quantify this pattern, Table 2.6 reports the widths of the 95% Wald confidence intervals for the heterogeneous model M4, expressed relative to the $N = 10,000$ population size. Values below one therefore indicate improved precision. The corresponding percentage reduction in interval width is shown in parentheses.

Table 2.6: Reduction in 95% Wald confidence interval widths with increasing population size for the heterogeneous model M4 in the τ -leaping analysis. Each entry reports the CI-width ratio relative to the $N = 10,000$ baseline; the corresponding percentage reduction is shown in parentheses. Values below one indicate improved precision.

Population	$\mathcal{R}_0$	ν	t_0	c^*
10,000	1.00 (0.00%)	1.00 (0.00%)	1.00 (0.00%)	1.00 (0.00%)
100,000	0.35 (65.34%)	0.50 (50.08%)	0.30 (70.46%)	0.39 (60.93%)
5,000,000	0.05 (94.95%)	0.07 (92.55%)	0.04 (95.88%)	0.06 (94.04%)

At the $N = 10,000$ scale, confidence intervals are wide, especially for ν and c^* . This reflects the additional uncertainty introduced by demographic process noise. Although the heterogeneous model M4 still gives estimates in the correct region, the fitted values $\hat{\nu} = 1.65$ and $\hat{c}^* = 0.339$ are imprecise relative to the data-generating values $\nu = 1.414$ and $c^* = 0.30$.

At $N = 100,000$, precision improves substantially. For M4, the confidence interval width for ν is reduced by about 50.08% relative to the $N = 10,000$ case, while the interval width for c^* is reduced by about 60.93%. Similar improvements occur for t_0 , with a reduction of 70.46%, and for $\mathcal{R}_0$, with a reduction of 65.34%. The median estimates also move closer to the data-generating values, with $\hat{\nu} = 1.49$ and $\hat{c}^* = 0.309$.

At $N = 5,000,000$, the stochastic results are close to the deterministic large-population limit. The confidence interval widths are reduced by more than 90% for all four fitted parameters relative to the $N = 10,000$ baseline. The median estimates are also very close to the data-generating

values: $\hat{\mathcal{R}}_0 = 3.00$, $\hat{\nu} = 1.42$, $\hat{t}_0 = 15.00$, and $\hat{c}^* = 0.300$. Thus, increasing population size reduces the effect of process noise and restores the precision expected from the deterministic likelihood analysis.

Importantly, the improvement in precision does not remove the qualitative identifiability issue. As shown by the joint distributions in Figure 2.12, the ν - c^* compensation direction remains visible under stochastic simulation, although the point clouds become tighter as N increases. Thus process noise mainly broadens the uncertainty around the same compensation geometry rather than creating a different inferential problem.

2.8 Epidemic trajectory forecasts

Quantifying parameter uncertainty is essential for assessing how much trust to place in model-based projections. Here we use the fitted homogeneous and heterogeneous models to generate predictive trajectories beyond the calibration window (days 1–100; black points). In the forecast window (days 101–200), we consider the scenario where NPIs are no longer enforced, so contacts return to baseline and we set $c(t) = 1$ for $t > 100$. For each model, we approximate the sampling distribution of the estimated transformed parameters by a multivariate normal (MVN) centred at the MLE and with covariance given by the inverse observed information. Drawing parameter sets from this MVN preserves the empirical correlation structure between parameters and allows us to propagate joint uncertainty into forward simulations. Figure 2.13 shows forecasts for **S3** (homogeneous population with NPIs in the calibration period). Both model classes are fitted on days 1–100 (black points) and then projected forward to day 200 with NPIs removed after day 100. Over the fitting window, the homogeneous and heterogeneous models follow the incidence curve almost identically. In the forecast window, the two median trajectories remain close: the heterogeneous fit produces a slightly lower peak than the homogeneous one, despite yielding a small positive estimate of the coefficient of variation $\hat{\nu} = 0.204$ (true value 0). The homogeneous model’s projection essentially coincides with the true generating trajectory (dashed black line), whereas the heterogeneous prediction lies just below it but with a noticeably wider predictive band. This increased spread is driven largely by uncertainty in coefficient of variation, ν , whose 95% confidence interval is relatively broad ($\hat{\nu}$ CI: 0.040–0.912).

As for the **S4**, shown in Figure 2.14, where the data arise from a heterogeneous model with NPIs in the calibration period, we observe divergence in the forecast window. Again, both models fit the observed days 1–100 visually well. However, once we project to day 200 with NPIs removed after day 100, the forecasts diverge substantially. The homogeneous model predicts a

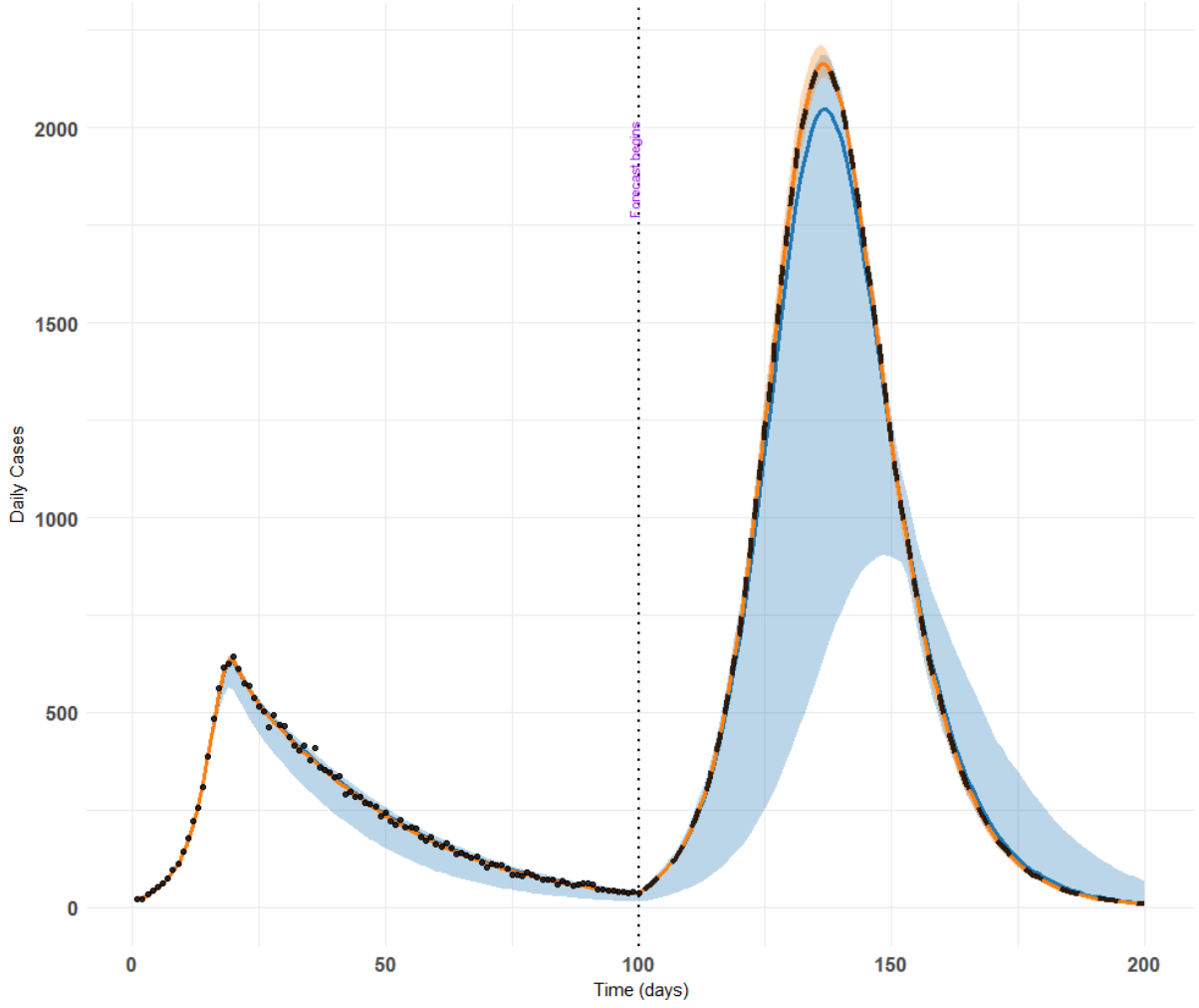


Figure 2.13: Forecast performance under homogeneous model (Scenario **S3**). Synthetic data are generated from a homogeneous population with NPIs and fitted on days 1–100 (black points); forecasts extend to day 200. The blue line and shaded band show the median and 95% predictive envelope from the heterogeneous fit (MVN parameter draws preserving the estimated correlation structure); the orange line is the homogeneous fit. The dashed black curve is the data-generating trajectory ($\mathcal{R}_0 = 3$, $t_0 = 15$, $c^* = 0.3$). Both models reproduce the observed wave and give similar forecasts; the heterogeneous band is wider, mainly due to the broad uncertainty around the small but non-zero estimate of ν .

very large second wave, with peak daily incidence around 2,750 cases—roughly four times the first-wave peak. The heterogeneous model instead produces a much smaller second wave, with most simulated peaks between 300 and 1,000 daily cases.

Comparing these projections with the true trajectory (dashed black line) reveals that the heterogeneous model closely reproduces both the timing and magnitude of the second wave, as expected because the data-generating trajectory was heterogeneous and the fitted estimate ($\hat{\nu} = 1.406$, 95% CI: 1.157–1.768) is close to the specified value. On the other hand, the homogeneous model, which does not account for heterogeneity, substantially overshoots the

prediction window as the NPIs are no longer enforced to "help" it.

These experiments highlight the forecasting implications of ignoring heterogeneity in susceptibility. In multi-wave settings, homogeneous models can substantially overestimate subsequent outbreaks, potentially leading to stringent or prolonged control measures.

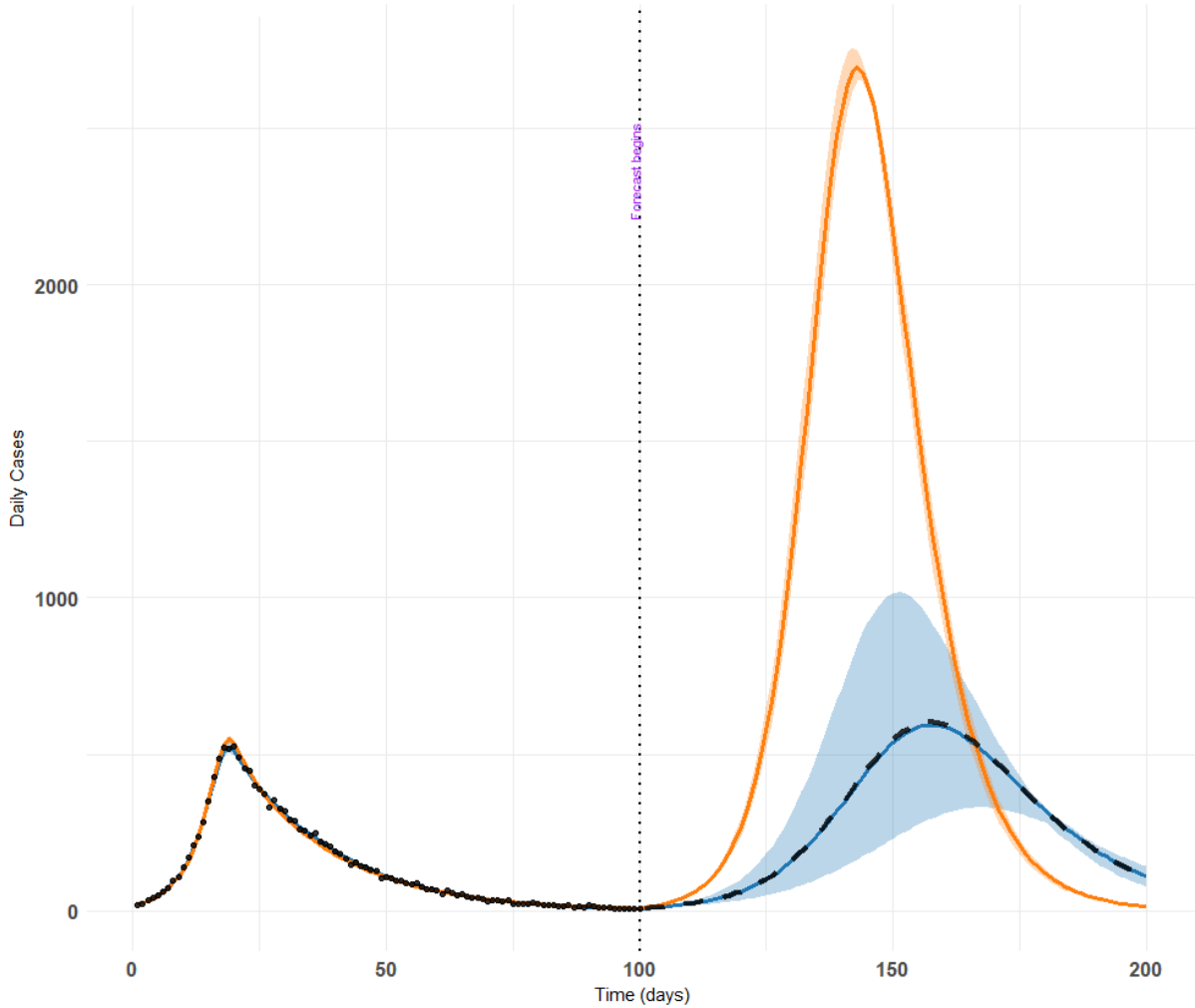


Figure 2.14: Forecast performance under heterogeneous model (Scenario **S4**). Data are generated from a heterogeneous population with NPIs and fitted on days 1–100 (black points); forecasts extend to day 200. The blue line and shaded band show the heterogeneous fit with its 95% predictive envelope; the orange line is the homogeneous fit. The dashed black curve is the true trajectory. The heterogeneous model tracks both waves with moderate uncertainty (driven by ν), whereas the homogeneous model severely overpredicts the second wave (peak $\sim 2,750$ daily cases), illustrating the forecasting bias that arises when fitting a homogeneous model to data generated under heterogeneous susceptibility.

2.9 Chapter Summary and Conclusions

In this chapter we examined the structural and practical identifiability of the parameter vector $(\mathcal{R}_0, \nu, t_0, c^*)$ for a heterogeneous-susceptibility SEIR model with NPIs using a *single* synthetic epidemic trajectory. The main result is that, although the model is structurally well-posed under

idealised assumptions, practical identifiability can be poor once heterogeneity in susceptibility and NPIs are both active in the same epidemic. In particular, the likelihood surface develops pronounced compensation ridges, with a negative dependence between $\mathcal{R}_0$ and t_0 , and a much stronger positive dependence between heterogeneity (ν) and intervention strength (c^*).

The chapter also shows that this ν - c^* confounding is the main practical difficulty in single-epidemic time series inference. When heterogeneity and intervention act together, markedly different combinations of ν and c^* can produce very similar incidence trajectories over the observed time window. This means that a good in-sample fit does not necessarily imply reliable recovery of the underlying mechanisms. In that sense, the problem is not simply whether the model fits, but whether the data contain enough distinct information to separate intrinsic population heterogeneity from changes in contact behaviour.

The main results in Chapter 2 were presented for the baseline intervention level $c^* = 0.30$, which represents a reasonable sustained NPI setting in the simulation design: with $\mathcal{R}_0 = 3$, the corresponding post-intervention controlled level is $3 \times 0.30 = 0.9 < 1$, so transmission is brought below the epidemic threshold once the intervention is established. Appendix C then checks that this baseline choice does not drive the broader conclusions by repeating the comparison under $c^* = 0.20$ and $c^* = 0.40$, corresponding to stricter and milder sustained control, respectively of intervention strengths.

The appendices strengthen this conclusion in three additional ways. Appendix B shows that the strong positive ν - c^* association is not tied to one fitting method or one choice of NPI profile. It persists across maximum likelihood, least squares, and Bayesian fits, and across linear, exponential, and logistic intervention functions. Section 2.7 then shows that the same compensation persists under fully stochastic tau-leaping simulations. In large populations, the tau-leaping results are close to the deterministic ODE analysis, while in smaller populations process noise mainly increases uncertainty rather than removing the ridge. This motivates the next step of the thesis. In Chapter 3, we ask whether joint analysis of more than one epidemic can provide the missing contrast that a single epidemic lacks, weaken the ν - c^* compensation ridge, and improve practical identifiability of the same shared parameter vector.

Chapter 3

Joint-Epidemic Inference and Parameter Identifiability

3.1 Introduction

In Chapter 2 we investigated the practical identifiability of the heterogeneous-susceptibility SEIR model with non-pharmaceutical interventions (NPIs) using a *single* epidemic wave. Working with noisy incidence data, we showed that the four parameters of interest - the basic reproduction number $\mathcal{R}_0$, the coefficient of variation in susceptibility ν , the onset timing of behavioural change t_0 , and the NPI floor c^* - are all, in principle, recoverable when either heterogeneity or interventions act in isolation. However, once susceptibility heterogeneity and NPIs act together, the likelihood surface develops pronounced compensation ridges: increases in ν can be offset by stronger or weaker interventions, and changes in $\mathcal{R}_0$ can be traded against shifts in t_0 .

The most severe of these ridges is the strong positive correlation between ν and c^* and a strong negative correlation between $\mathcal{R}_0$ and t_0 . Chapter 2 demonstrated that the ν - c^* correlation is a robust feature of single-time series inference. It persists across different NPI parameterisations and across statistical frameworks (see appendix B for details). In other words, the difficulty is not a consequence of a particular fitting method or modelling the NPI, but a structural limitation of trying to separate population heterogeneity from intervention strength from a single epidemic time series.

Our intuition for the central idea of this chapter is that multiple epidemics can act as counterfactuals for one another. A single epidemic may not allow the effects of susceptibility heterogeneity, ν , and sustained intervention strength, c^* , to be separated reliably. In the framework considered here, the epidemics share $(\mathcal{R}_0, \nu, t_0, c^*)$ and the same intervention profile,

but begin from different initial conditions. They therefore experience different susceptible-pool depletion histories, providing variation that is absent from a single trajectory.

This intuition follows from the compensation argument in Section 1.8 and Section 2.3.2. From equation (2.8), epidemic j may primarily constrain

$$c_{\text{eff},j} \approx c^* \left(\frac{S_j^*}{N_j} \right)^{\nu^2}, \quad j = 1, 2.$$

Taking logarithms and subtracting the two relationships eliminates the shared intervention level:

$$\log c_{\text{eff},1} - \log c_{\text{eff},2} \approx \nu^2 \left[\log \left(\frac{S_1^*}{N_1} \right) - \log \left(\frac{S_2^*}{N_2} \right) \right]. \quad (3.1)$$

The contrast is informative about ν only to the extent that the two epidemics generate distinct depletion histories. If their representative susceptible fractions are equal or nearly equal, the second trajectory adds little information in this direction. Joint fitting therefore improves separation not simply because two epidemics are observed, but because the same shared parameters must account for trajectories generated from different initial conditions. A third epidemic may add further independent variation and improve precision, but does not introduce a fundamentally different separation mechanism. This argument provides conceptual motivation rather than a formal proof of identifiability.

The counterfactual interpretation also connects with empirical joint analyses such as Gomes et al. [4], in which related epidemics were analysed together to investigate how individual variation in susceptibility or exposure affects estimates of epidemic thresholds.

In this chapter we formalise this idea by treating multiple epidemics as approximate counterfactuals of one another and asking how much identifiability can be recovered when such trajectories are analysed jointly.

Motivated from Chapter 2, the aims of this chapter are threefold:

1. **Compare likelihood geometry under single- and multi-epidemic inference.** We compare the likelihood geometry under both single epidemic and joint two-epidemic analysis in which the epidemics share $(\mathcal{R}_0, \nu, t_0, c^*)$ but differ in initial conditions. Using the observed information matrix (eigenvalues/eigenvectors), profile-likelihood confidence intervals, and likelihood-surface contour plots, we quantify how moving from one to two epidemics changes curvature, numerical conditioning, and parameter correlations, with particular focus on the ν - c^* compensation ridge.

2. **Test limits within the likelihood framework.** We evaluate diminishing returns when adding further epidemics (two vs. three) and summarise when multi-epidemic sharing meaningfully improves practical identifiability.
3. **Test limits in higher-dimensional parameter spaces.** We examine how far the precision gains from joint multi-epidemic inference can be maintained when selected sharing assumptions are relaxed. Specifically, we consider higher-dimensional extensions in which some parameters are allowed to be epidemic-specific (e.g. epidemic-specific onset times and/or reproduction numbers), and assess whether this added flexibility reintroduces compensation effects and widens uncertainty intervals.

This chapter is organised as follows. Section 3.2 summarises the heterogeneous-susceptibility SEIR model and the NPI parametrisation used. Section 3.3 introduces the multi-epidemic likelihood and the information-based diagnostics used to quantify identifiability gains. Section 3.4 presents the main single- versus two-epidemic results under the baseline design. Section 3.5 extends the framework to a three-epidemic setting to evaluate diminishing returns from adding further trajectories. Lastly, the chapter summarises and concludes in Section 3.6. Supplementary sensitivity results on intervention strength and initial conditions are given in Appendix C, while Bayesian higher-dimensional extensions are reported in Appendix D, where Bayesian inference is used because optimisation-based MLE can become numerically unstable as the parameter dimension increases.

3.2 Mathematical Model and NPIs

This chapter uses the same reduced heterogeneous-susceptibility SEIR model and the same piecewise linear NPI ramp introduced in Chapter 2 (see Section 2.2 and equation (2.3)). We therefore do not repeat the governing equations here.

We consider J epidemic trajectories indexed by $j = 1, \dots, J$. For each epidemic i , the state variables are $(S_j(t), E_j(t), I_j(t), R_j(t))$ in a population of size N_j . In the baseline two- and three-epidemic experiments (Sections 3.4 and 3.5), we keep the NPI schedule identical across epidemics, $t_0^{(1)} = \dots = t_0$ and $c_j(t) \equiv c(t)$, so that t_1 as well as the parameters $(\mathcal{R}_0, \nu, t_0, c^*)$ are shared. Differences between epidemics in these designs arise only through the initial conditions $(E_{0,j}, I_{0,j})$, producing trajectories with distinct depletion histories under the same policy.

3.3 Methodology: The Joint-Epidemic Framework

Here we formally extend the single-epidemic likelihood framework to analyse multiple concurrent epidemic trajectories. This section defines the joint likelihood for shared parameters, the curvature-based diagnostics (Hessian eigenvalues and condition number), and the simulation design used to compare $J = 1$ versus $J = 2$ inference.

3.3.1 Joint Likelihood Function

We observe two epidemic trajectories (e.g., two regions), indexed by $j \in \{1, 2\}$. For epidemic j , the data are daily incidence counts $\mathbf{y}_j = (y_{j,1}, \dots, y_{j,T})$, where $t = 1, \dots, T$ denotes days.

The shared parameter vector for both epidemic is

$$\boldsymbol{\theta} = (\mathcal{R}_0, \nu, t_0, c^*).$$

For each epidemic j , with epidemic-specific initial conditions $(\mathbf{E}_{0,j}, \mathbf{l}_{0,j})$, will have a model-implied (from the system (2.1)) mean daily incidence sequence

$$\mathbf{x}_j(\boldsymbol{\theta}) = (x_{j,1}(\boldsymbol{\theta}), \dots, x_{j,T}(\boldsymbol{\theta})),$$

where $x_{j,t}(\boldsymbol{\theta})$ denotes the predicted mean incidence on day t .

Observation model. As in Section 2.5, we assume the observations $y_{j,t}$ are conditionally independent given the model states $x_{j,t}$. The joint likelihood factorises as

$$L(\boldsymbol{\theta} \mid \mathbf{y}_1, \mathbf{y}_2) = \prod_{t=1}^T f(y_{1,t} \mid x_{1,t}(\boldsymbol{\theta})) \prod_{t=1}^T f(y_{2,t} \mid x_{2,t}(\boldsymbol{\theta})), \quad (3.2)$$

where $f(\cdot)$ is the probability mass function (Poisson or Negative Binomial), and the joint log-likelihood is additive:

$$\ell(\boldsymbol{\theta}) := \log L(\boldsymbol{\theta} \mid \mathbf{y}_1, \mathbf{y}_2) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}), \quad \ell_j(\boldsymbol{\theta}) = \sum_{t=1}^T \log f(y_{j,t} \mid x_{j,t}(\boldsymbol{\theta})). \quad (3.3)$$

MLE. We obtain the maximum likelihood estimate by

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} := \arg \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}).$$

We compare $J = 1$ (single epidemic, fitting epidemic 1 only) against $J = 2$ (two epidemics fitted jointly using (3.2)), under the same data-generating grid in Section 3.3.4.

3.3.2 Profile Likelihood Analysis

Let

$$\mathbf{H} = -\nabla^2 \ell(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (3.4)$$

denote the observed information matrix, evaluated at the maximum-likelihood estimate. When $\mathbf{H}$ is positive definite, the diagonal elements of $\mathbf{H}^{-1}$ provide the approximate parameter variances on the transformed scale, and their square roots are the Hessian-based standard errors.

We use profile likelihoods to assess practical identifiability of each component of $\boldsymbol{\theta} = (\mathcal{R}_0, \nu, t_0, c^*)$ for both the single-epidemic ($J = 1$) and two-epidemic ($J = 2$) likelihoods in (3.3). Let $\boldsymbol{\theta} = (\psi, \boldsymbol{\phi})$, where the scalar parameter of interest is $\psi \in \{\mathcal{R}_0, \nu, t_0, c\}$ and $\boldsymbol{\phi}$ collects the remaining three parameters. The profile log-likelihood for ψ is

$$\ell_{\text{prof}}(\psi) = \max_{\boldsymbol{\phi}} \ell(\psi, \boldsymbol{\phi}). \quad (3.5)$$

Let $\hat{\boldsymbol{\theta}}$ denote the MLE and $\ell_{\text{max}} = \ell(\hat{\boldsymbol{\theta}})$. The likelihood-ratio statistic is

$$\Lambda(\psi) = 2[\ell_{\text{max}} - \ell_{\text{prof}}(\psi)]. \quad (3.6)$$

We use the 95% cutoff $\chi_{0.95,1}^2 \approx 3.84$, equivalently $\ell_{\text{prof}}(\psi) \geq \ell_{\text{max}} - 1.92$, to construct profile-likelihood confidence intervals. In the results we compare these intervals between $J = 1$ and $J = 2$, analysing how the ν - c^* ridge affects the profiles and how joint inference sharpens curvature.

In our implementation, the profile was computed on a grid of candidate values for ψ , centred on $\hat{\psi}_{\text{MLE}}$ and spanning ± 5 Hessian-based standard errors on the transformed scale. At each grid point, ψ was held fixed at its transformed value and the remaining parameters $\boldsymbol{\phi}$ were re-optimised using the Nelder–Mead algorithm in `optim()` (maximum 1 000 iterations), with the same log/logit transformations described in Section 2.5. The likelihood-ratio statistic $\Lambda(\psi)$ was evaluated at every grid point and the 95% profile-likelihood confidence interval was taken as the range of ψ values satisfying $\Lambda(\psi) \leq 3.84$. This procedure was applied identically for $J = 1$ and $J = 2$ fits to enable direct comparison of interval widths.

3.3.3 Eigenvalue and Hessian analysis

Because the joint log-likelihood is additive,

$$\ell(\boldsymbol{\theta}) = \sum_{j=1}^J \ell_j(\boldsymbol{\theta}),$$

the observed information matrix in equation (3.4) also aggregates additively:

$$\mathbf{H} = \sum_{j=1}^J \mathbf{H}_j.$$

We analyse the eigen-decomposition

$$\mathbf{H} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T, \tag{3.7}$$

where

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4), \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 > 0,$$

and $\mathbf{V}$ contains the associated eigenvectors. Large eigenvalues correspond to directions of high curvature, representing better-constrained combinations of $\mathcal{R}_0$, ν , t_0 , and c^* , whereas small eigenvalues indicate weakly constrained directions that align with compensation ridges.

We summarise numerical stability using the condition number

$$\kappa = \frac{\lambda_1}{\lambda_4}.$$

A large value of κ indicates an ill-conditioned likelihood surface. We compute κ and the Hessian-based correlation structure for both the single-epidemic ($J = 1$) and two-epidemic ($J = 2$) fits.

3.3.4 Simulation Design: The Large vs. Small Epidemic

The baseline simulation grid compares single-epidemic inference against a two-epidemic joint inference. The shared data-generating parameters are given in Table 2.1.

The only differences are the distinct initial seed sizes :

1. **Epidemic 1 (Large):** $E_0 = 1000$, $I_0 = 400$, representing substantial early spread.
2. **Epidemic 2 (Small):** $E_0 = 100$, $I_0 = 40$, representing limited early spread (Figure 3.1).

The baseline pair $(I_0^{(1)}, I_0^{(2)}) = (400, 40)$ was chosen to create a clear but still controlled contrast

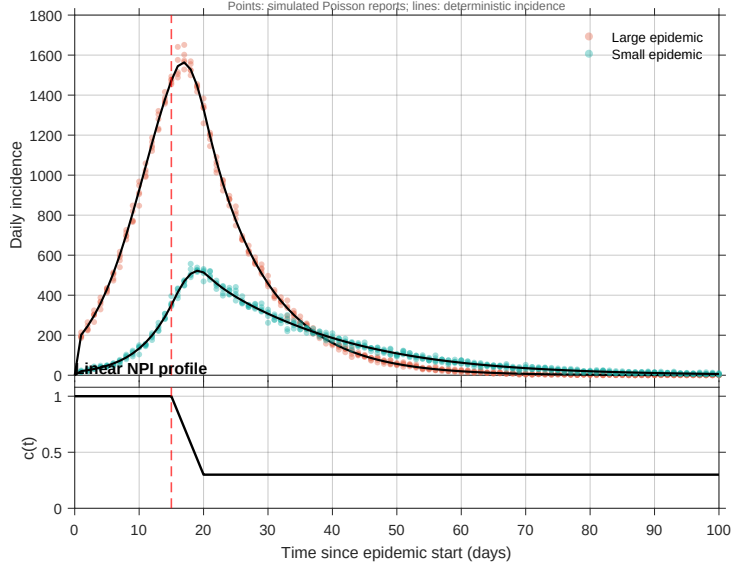


Figure 3.1: Baseline two-epidemic simulation design used for the joint-inference experiment. The upper panel shows daily incidence for the Large and Small epidemics. Coloured points are Poisson-simulated daily reports, while black curves show the corresponding deterministic incidence trajectories. The lower panel shows the shared piecewise-linear NPI profile $c(t)$: transmission remains at baseline before behavioural change begins at $t_0 = 15$ days, decreases linearly until $t_1 = 20$ days, and is then held at the sustained level $c^* = 0.3$ until the end of the fitting window. The two epidemics share $(\mathcal{R}_0, \nu, t_0, c^*)$ and differ only in their initial conditions, generating different depletion histories under the same intervention schedule.

in epidemic stage under the same shared parameter vector. The smaller seed ($I_0 = 40$) enables the single-epidemic to have sufficient initial start to retain a pronounced early-growth phase, while the larger seed ($I_0 = 400$) places the companion epidemic substantially further along the trajectory before the intervention takes effect. This produces visibly different depletion histories without changing $(\mathcal{R}_0, \nu, t_0, c^*)$, which is exactly the type of contrast needed to test whether joint fitting can constrain the ν - c^* compensation ridge. We simulated 200 stochastic datasets¹ under this. The Parameters $(\mathcal{R}_0, \nu, t_0, c^*)$ were re-estimated using maximum likelihood estimation described in section 3.3.1.

3.4 Results: Profile likelihood and Hessian Analysis single vs. Two Epidemics

We first discuss the results of the profile likelihood estimation and Hessian eigenvalue analysis for all the 200 synthetic datasets jointly for both single-epidemic and two-epidemic configurations

¹A sample size of 200 replicated datasets was chosen to balance Monte Carlo precision against computational cost. By the central limit theorem, the Monte Carlo standard error of a replicated-sample mean is $s/\sqrt{200}$, where s is the empirical standard deviation of the quantity being summarised, so the simulation error in reported means is about 0.071s. For estimated coverage probabilities, the corresponding standard error is $\sqrt{\hat{p}(1-\hat{p})/200}$, which is at most 0.035 and is about 0.015 when $\hat{p} \approx 0.95$. Thus 200 replicates provide adequate precision for the distributional summaries reported here while remaining computationally feasible, since each replicate requires numerical ODE integration and iterative likelihood optimisation.

and performed. We start with bias and precision gains.

3.4.1 Quantitative Improvements: Precision and Bias

The transition from single-epidemic to two-epidemic analysis yielded substantial improvements in parameter precision across all four shared parameters. Table 3.1 summarizes the mean parameter estimates and confidence interval (CI) widths derived from the profile likelihoods.

Table 3.1: Parameter estimates and confidence intervals from single vs two epidemics (NPI $c^* = 0.30$).

Parameter	Single Epidemic		Two Epidemics		Coverage Probability (%)	
	Mean (SD)	CI Width	Mean (SD)	CI Width	Single	Two
$\mathcal{R}_0$	3.00 (0.031)	0.098	3.00 (0.011)	0.040	87.0	92.0
ν	1.417 (0.245)	0.666	1.414 (0.011)	0.039	85.5	95.5
t_0	14.95 (0.487)	1.569	15.01 (0.209)	0.706	87.5	93.0
c^*	0.30 (0.015)	0.042	0.30 (0.002)	0.007	83.5	92.5

The highest improvement occurred for the coefficient of variation (ν). In the single-epidemic case, the profile likelihood for ν yielded a broader confidence interval of width 0.666 when compared to the two-epidemic with a confidence interval width of only 0.039. This corresponds to a reduction of approximately 94% in interval width, or an ≈ 17 -fold gain in precision.

Substantial gains were also observed for the intervention and transmission parameters. For the intervention strength c^* , the confidence interval width shrank on average from 0.042 (single epidemic) to 0.007 (two epidemics), an 83% reduction (≈ 6 -fold narrower). The timing parameter t_0 shows a 55% reduction in interval width (1.569 to 0.706, ≈ 2.2 -fold narrower), while the basic reproduction number $\mathcal{R}_0$ improves by about 59% (0.098 to 0.040, ≈ 2.5 -fold narrower). These reductions in uncertainty are reflected in the higher coverage probabilities for the two-epidemic fits in Table 3.1. The experiment in section 3.3.4 is repeated for using two more NPI levels $c^* = 0.2$ and 0.4 as sensitivity checks and reported in supplementary results Appendix C together with the baseline case $c^* = 0.30$.

3.4.2 Numerical Stability and Condition Number

To assess numerical stability, we analysed the condition numbers of the Hessian matrices at the maximum likelihood estimates. Table 3.2 as well as Figure C.1 in Appendix C summarize these results along with two other values of $c^* = 0.20$ and $c^* = 0.40$.

The two-epidemic approach has improved numerical stability, with a median condition number of 59.3 compared to 3069.6 for the single-epidemic approach. All the condition numbers in the

Table 3.2: Hessian Matrix Condition Numbers Comparison

Metric	Single epidemic	Two epidemics
Mean	4347.5	59.2
Median	3069.6	59.3
Minimum	376.5	47.7
Maximum	40911.3	69.3
Standard deviation	4703.0	3.2
Relative improvement	73.43×	

single-epidemic approach are very large as seen in Table 3.2, parameters become increasingly difficult to estimate accurately. Combined with the narrower confidence intervals in Section 3.4.1, and compensation ridge in Figure 3.3 in section 3.4.4 we infer that the two-epidemic joint fits improves identifiability of the shared parameter vector $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$.

3.4.3 Parameter Correlations

The primary purpose of this comparison is to assess whether using epidemics at contrasting stages of development reduces the compensation directions seen in single-epidemic fitting. We summarise the local correlation structure in Table 3.3, using correlations derived from the average inverse Hessian matrix across the replicated fits which can be visualised from the joint-distributions of key parameters in Figure 3.2 of the replicated profile-MLE estimates.

Table 3.3: Parameter correlations derived from the average inverse Hessian matrix across 200 simulated datasets for the baseline setting $c^* = 0.3$. The single-epidemic fit uses one epidemic trajectory, whereas the two-epidemic fit jointly fits the large and small epidemics with shared $(\mathcal{R}_0, \nu, t_0, c^*)$.

Parameter pair	Single epidemic	Two epidemics
$\mathcal{R}_0-t_0$	-0.70	-0.68
$\mathcal{R}_0-\nu$	0.65	-0.04
$\mathcal{R}_0-c^*$	0.55	-0.39
$t_0-\nu$	-0.05	0.39
t_0-c^*	0.08	0.12
$\nu-c^*$	0.94	0.39

Table 3.3 shows that the dominant single-epidemic dependence is the positive $\nu-c^*$ correlation (0.94), reflecting the compensation between susceptibility heterogeneity and intervention strength. Under the two-epidemic design this correlation is reduced to 0.39, indicating that the second trajectory helps constrain the main compensation direction. The correlations involving $\mathcal{R}_0$ also change substantially: the $\mathcal{R}_0-\nu$ correlation falls from 0.65 to -0.04, while the $\mathcal{R}_0-c^*$ association changes from 0.55 to -0.39. The $\mathcal{R}_0-t_0$ correlation remains strongly negative in both designs,

suggesting that transmission strength and intervention timing continue to trade off locally even after adding the second epidemic.

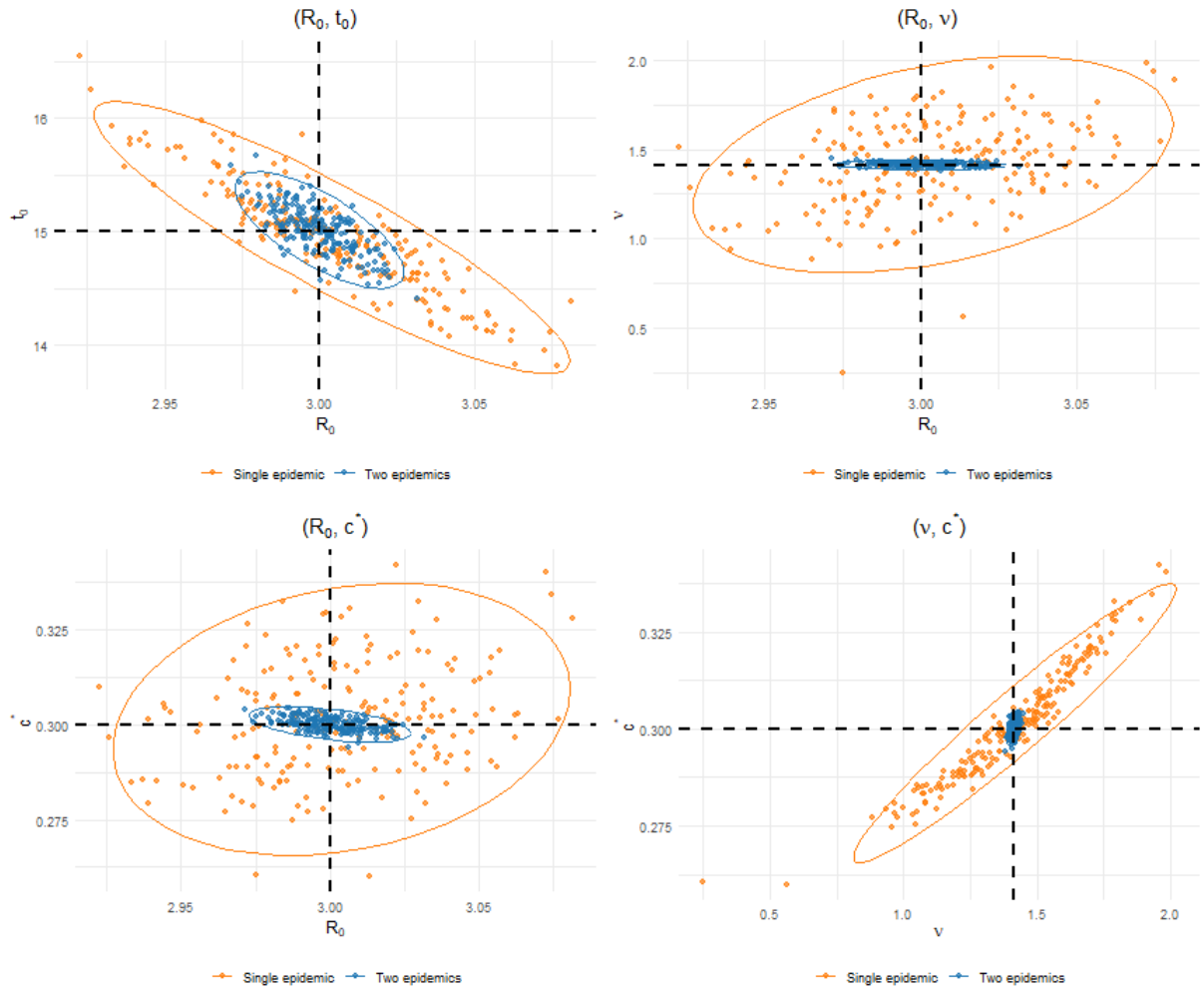


Figure 3.2: Joint distributions of the profile-likelihood *estimates* across the 200 replicate datasets, overlaying the single-epidemic (orange) and two-epidemic (blue) fitting designs, for the scenario with $c^* = 0.3$. Panels: (R_0, t_0) , (R_0, ν) , (R_0, c^*) , and (ν, c^*) . Each point is one replicate’s estimate; solid curves are per-design normal-approximation 95% contours; dashed black lines mark the true values ($R_0 = 3, \nu = 1.414, t_0 = 15, c^* = 0.3$). The two-epidemic estimates form a substantially tighter cloud than the single-epidemic estimates, and the elongated single-epidemic spread along the (ν, c^*) direction is visibly contracted under the two-epidemic design.

3.4.4 Visualising the compensation ridge: bivariate likelihood surfaces and profile likelihoods

To visualize the geometric structure of the ν - c^* compensation ridge, we constructed bivariate likelihood surfaces over a grid of (ν, c^*) values while holding R_0 and t_0 fixed at their maximum likelihood estimates. For each grid point, we computed the Poisson log-likelihood by integrating the SEIR model forward and comparing predicted to observed cases. The grid was centered on

the MLE with a range of approximately $\pm 15\%$ for each parameter, providing sufficient resolution to capture the local likelihood geometry without excessive computation.

Figure 3.3 shows the resulting surfaces for a representative dataset generated with $c^* = 0.3$, $\nu = 1.414$, $\mathcal{R}_0 = 3.0$, and $t_0 = 15$ days. The left panel shows the single-epidemic surface; the right panel shows the joint two-epidemic surface. Filled contours indicate relative log-likelihood, with darker regions (purple) near the maximum and lighter regions (yellow) where likelihood approaches zero. The red cross marks the true parameter values.

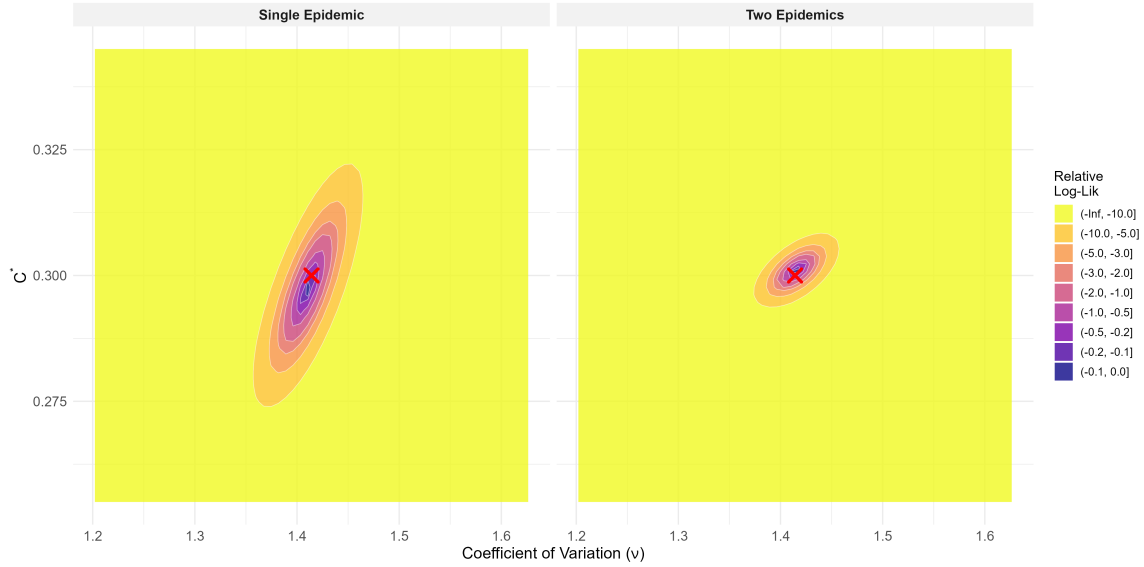


Figure 3.3: Bivariate likelihood surfaces for the (ν, c^*) parameter space. **Left:** single epidemic. **Right:** two epidemics fitted jointly. Filled contours show relative log-likelihood from maximum (dark purple) to near-zero (yellow). The red cross marks the true values ($\nu = 1.414$, $c^* = 0.3$). The single-epidemic surface exhibits pronounced elongation along a diagonal ridge, indicating that many (ν, c^*) pairs produce similar likelihood. The two-epidemic surface is markedly more circular and concentrated, indicating that the ridge is broken when fitting both epidemics jointly.

The single-epidemic surface shows pronounced elongation along a diagonal ridge running from lower-left (low ν , low c^*) to upper-right (high ν , high c^*). This elongation is the visual manifestation of the ν - c^* compensation: moving along the ridge (simultaneously increasing both heterogeneity and NPI floor, or simultaneously decreasing both) maintains high likelihood because the two effects partially offset one another. The positive slope of the ridge corresponds to the strong positive correlation (≈ 0.97) between ν and c^* in the single-epidemic analysis. The extended colour gradient along this ridge indicates that many parameter combinations remain plausible given the data.

The two-epidemic surface is markedly different. The contours are more circular and concentrated around the true values, with a steeper drop-off in all directions. This change in geometry reflects the fact that the two epidemics provide constraints along different directions in

parameter space: the second epidemic constrains the compensation ridge through its differing initial conditions and trajectory shape. The correlation between ν and c^* drops to approximately 0.32 in the two-epidemic case, consistent with the more symmetric, circular contour structure.

To make this ridge-breaking effect easier to compare directly, we also show the corresponding contour overlay in Figure 3.4. This figure plots selected relative log-likelihood contours for the single- and two-epidemic fits on the same (ν, c^*) axes. It is therefore a likelihood ridge map: it summarises the same local likelihood geometry as contours and makes the contraction of the ridge visually explicit.

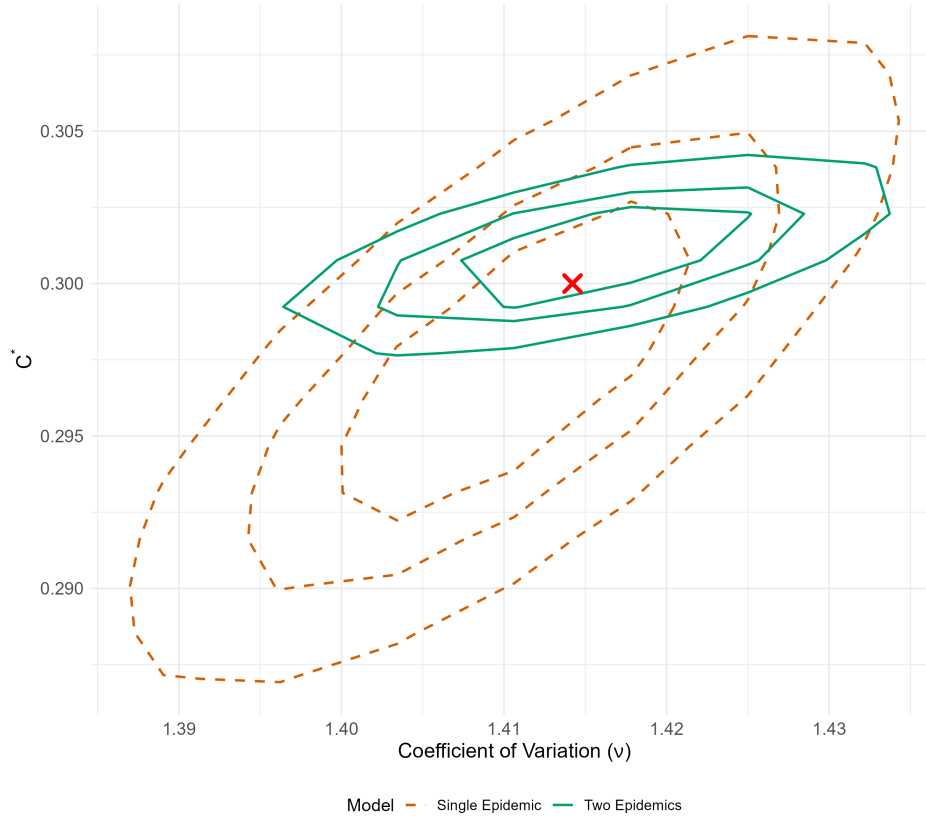
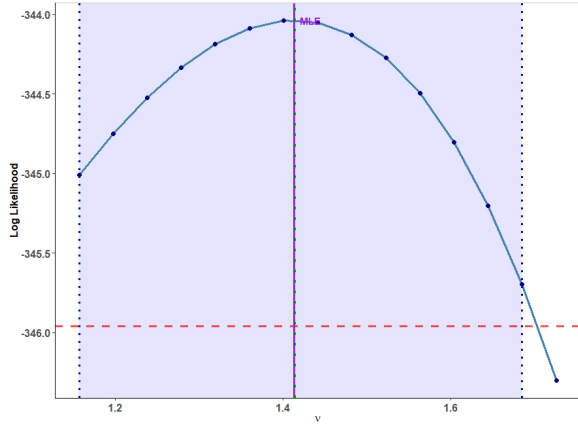


Figure 3.4: Overlay of bivariate likelihood contours in the (ν, c^*) parameter plane for the baseline intervention setting $c^* = 0.3$. Orange dashed contours show the single-epidemic fit and green solid contours show the joint two-epidemic fit. Contours correspond to relative log-likelihood drops of -0.5 , -1 , and -2 from the maximum. The red cross marks the data-generating values $(\nu, c^*) = (1.414, 0.3)$. The single-epidemic contours are elongated along a diagonal ridge, whereas the two-epidemic contours are tighter and more concentrated.

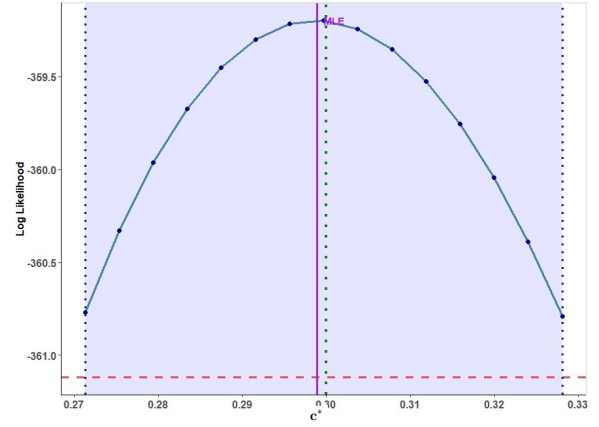
The baseline contour overlay in Figure 3.4 is extended in Appendix C, where the same comparison is repeated for $c^* = 0.20$ and $c^* = 0.40$. Those sensitivity plots show that the single-epidemic ridge persists across intervention strengths, while the degree of ridge-breaking under joint inference depends on the strength of control.

As a complementary one-dimensional diagnostic, Figure 3.5 shows profile likelihoods for the two parameters that define the main ν - c^* compensation ridge. The profile likelihood varies one

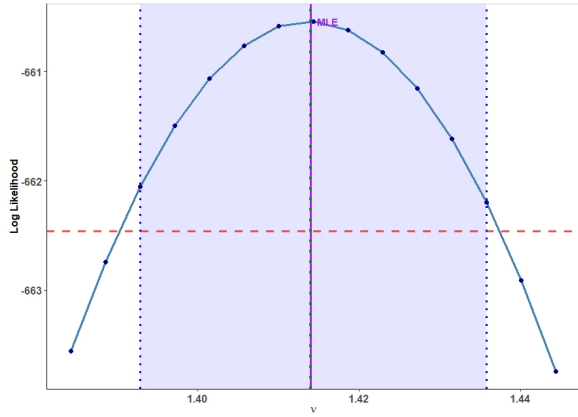
parameter at a time and re-optimises over the remaining fitted parameters.



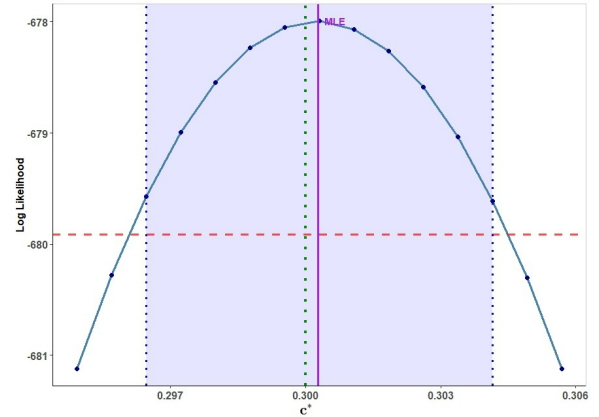
(a) ν , single epidemic.



(b) c^* , single epidemic.



(c) ν , two epidemics.



(d) c^* , two epidemics.

Figure 3.5: Profile likelihood curves for the coefficient of variation ν and intervention floor c^* in a representative baseline dataset with $c^* = 0.30$. Horizontal dashed red lines indicate the 95% profile-likelihood threshold; green dotted vertical lines mark the data-generating values, and purple vertical lines mark the MLEs. The two-epidemic profiles are sharper, consistent with the shorter 95% confidence-interval widths reported in Table 3.1.

The profile likelihoods reinforce the message from the two-dimensional surfaces. In the single-epidemic fit, the ν and c^* profiles are relatively broad, indicating that both ridge parameters are weakly determined when other parameters can compensate. In the two-epidemic fit, both profiles are much sharper around the data-generating values. Thus, the bivariate likelihood surfaces, contour overlay, and profile likelihoods all support the same conclusion: joint two-epidemic fitting reduces the practical non-identifiability associated with the ν – c^* compensation ridge.

3.5 Testing Limits: The Three-Epidemic Extension

Having established the gains from fitting two epidemics jointly, we next ask whether adding a third trajectory yields further improvement in identifiability, or whether the returns diminish as additional datasets provide overlapping information. To investigate this, we extend the framework

to $J = 3$ epidemics.

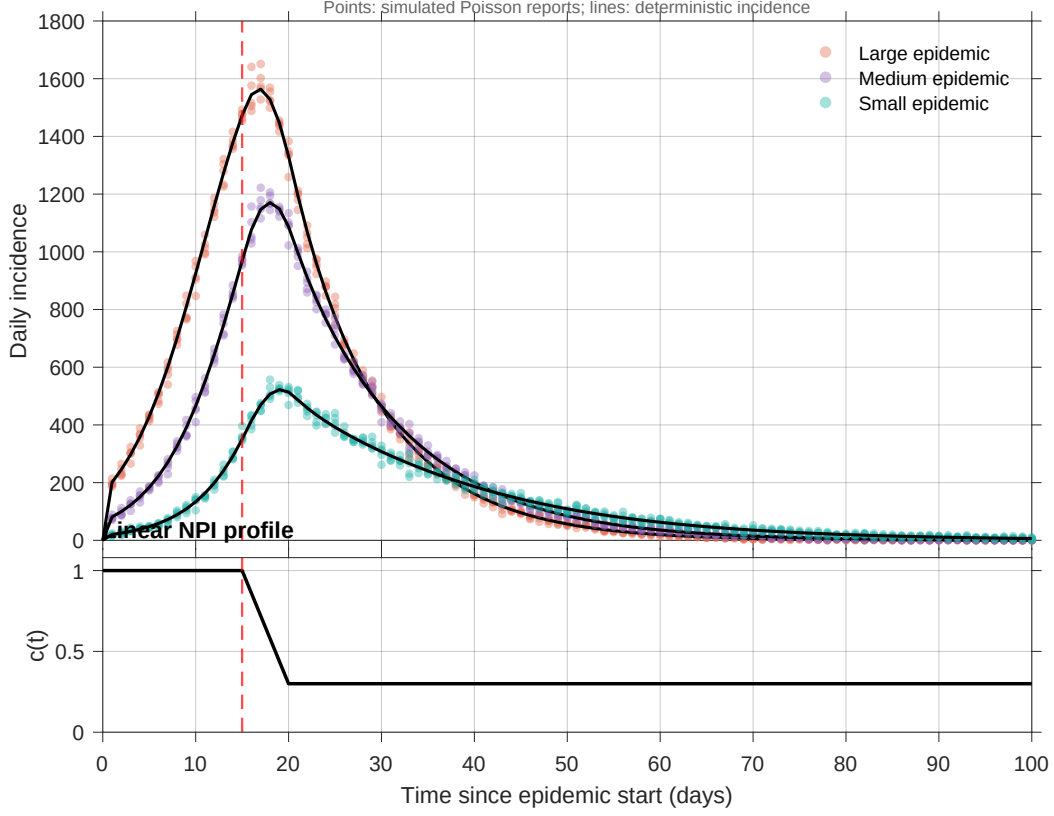


Figure 3.6: Simulation design for the three-epidemic extension. The upper panel shows Large, Medium and Small epidemic trajectories generated under the same shared parameter vector and the same linear NPI profile. Coloured points are Poisson-simulated daily reports and black curves are the corresponding deterministic incidence trajectories. The lower panel shows the shared contact multiplier $c(t)$, with behavioural change beginning at $t_0 = 15$ days, a linear reduction to $c^* = 0.3$ by $t_1 = 20$ days, and a return to $c(t) = 1$ at the end of the simulated window. The Medium trajectory is added to test whether an intermediate depletion history gives substantial extra information beyond the Large–Small pair.

3.5.1 Simulation Design

We augment the baseline “Large” and “Small” design with an intermediate “Medium” epidemic. The three trajectories differ only through their initial conditions, while sharing the same transmission and NPI parameters $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$.

- **Large epidemic:** $I_0 = 400$, $E_0 = 1000$.
- **Medium epidemic:** $I_0 = 160$, $E_0 = 400$.
- **Small epidemic:** $I_0 = 40$, $E_0 = 100$.

All three epidemics share the identical data-generating parameters used in the previous section ($\mathcal{R}_0 = 3.0$, $\nu = 1.414$, $t_0 = 15$, $c^* = 0.3$) (see example in Figure 3.6). As in Section 3.3.1,

we assume independent Poisson observation models with model-implied mean incidence $\mathbf{x}_j(\boldsymbol{\theta})$ for epidemic $j \in \{1, 2, 3\}$ and day t . The joint likelihood for $J = 3$ is therefore

$$L(\boldsymbol{\theta} \mid \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) = \prod_{j=1}^3 \prod_{t=1}^T f(y_{j,t} \mid x_{j,t}(\boldsymbol{\theta})), \quad \ell(\boldsymbol{\theta}) = \sum_{j=1}^3 \ell_j(\boldsymbol{\theta}), \quad (3.8)$$

and $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ is obtained by maximising $\ell(\boldsymbol{\theta})$. The optimisation and profile-likelihood procedures are identical to those described in Sections 2.5 and 3.3.2, extended to sum over three rather than two individual log-likelihoods.

3.5.2 Results: Diminishing Returns Two vs. Three Epidemics

We now briefly report the results of the extended analysis to three epidemics to assess whether additional data yield substantial further gains in precision.

Table 3.4 summarises the mean parameter estimates, average profile-likelihood confidence interval (CI) widths, and coverage probabilities for the two and three-epidemic fits. Adding a third epidemic leads to only modest reductions in CI width. For the basic reproduction number $\mathcal{R}_0$ and onset of behavioural change t_0 , the average widths shrink by around 17% ($\approx 1.2\times$ narrower) and 22% ($\approx 1.3\times$ narrower), respectively relative to two epidemics. The effects on the coefficient of variation (CV) and intervention strength c^* are even smaller, with CI width reductions of only about 7% and 6% (roughly $1.1\times$ narrower). These increments are insignificant compared with the large improvements obtained when moving from one to two epidemics of similar magnitude.

Coverage probabilities remain close to the nominal 95% level for both designs, with only small fluctuations between the two- and three-epidemic configurations. For $\mathcal{R}_0$ and t_0 the third epidemic slightly increases coverage, whereas for ν and c^* coverage decreases by a couple of percentage points; these differences are small relative to the sampling variation across the 200 simulated datasets and do not suggest any systematic gain or loss of reliability.

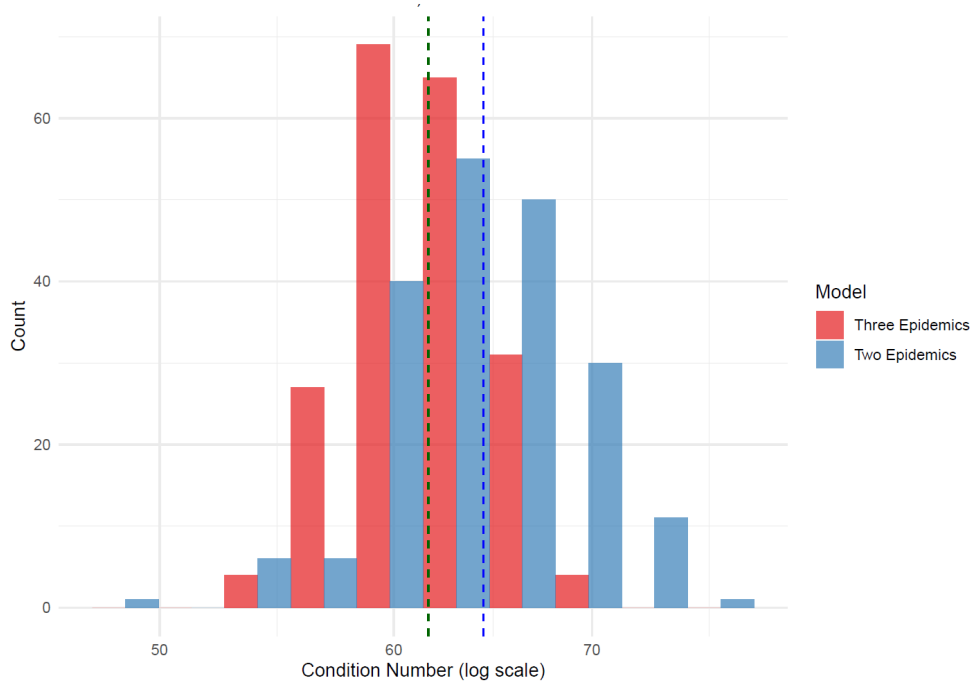


Figure 3.7: Distribution of Hessian condition numbers for two-epidemic (blue) vs. three-epidemic (red) fits across 200 simulated datasets ($c^* = 0.3$). Both designs produce well-conditioned Hessians with condition numbers concentrated between roughly 50 and 75. The three-epidemic distribution is only slightly left-shifted relative to the two-epidemic case, confirming that the main gains in numerical stability are already realised with two epidemics.

Again on whether the third epidemic improved the curvature of the likelihood surface, a similar pattern emerges when examining numerical stability via the average condition numbers. Figure 3.7 compares the distributions of Hessian condition numbers for the two- and three-epidemic fits across the 200 datasets. Both designs yield condition numbers tightly concentrated in the 50–75 range, with means in the low 60s. The three-epidemic distribution is slightly shifted to the left, indicating a small improvement in curvature, but the overlap between the histograms is substantial. In contrast to the higher reduction in condition numbers observed when moving from one to two epidemics, the third epidemic offers only marginal additional gains in conditioning.

The correlation comparison in Table 3.5 supports the same interpretation. The strongest dependence remains the negative $\mathcal{R}_0$ – t_0 association, and the ν – c^* correlation is not reduced by adding the medium epidemic. Thus, while the estimates remain accurate across both models, the gains in precision are marginal. These results support the interpretation that two epidemics are practically sufficient in this simulation design: they alleviate the key compensation mechanisms, deliver near-nominal coverage, and yield well-conditioned likelihood surfaces. A third epidemic provides only incremental refinements in precision and stability, with rapid diminishing returns

relative to the added modelling complexity.

Table 3.4: Parameter estimates and confidence intervals from two vs. three epidemics (NPI $c^* = 0.30$).

Parameter	Two Epidemics		Three Epidemics		Coverage (%)	
	Mean (SD)	CI Width	Mean (SD)	CI Width	Two	Three
$\mathcal{R}_0$	3.000 (0.011)	0.040	3.000 (0.009)	0.033	92.0	94.5
ν	1.415 (0.011)	0.039	1.414 (0.010)	0.036	95.5	92.5
t_0	15.013 (0.209)	0.706	14.997 (0.145)	0.552	93.0	94.5
c^*	0.300 (0.002)	0.007	0.300 (0.002)	0.006	92.5	90.5

Table 3.5: Parameter correlations derived from the average inverse Hessian matrix for the two- and three-epidemic designs. The two-epidemic design jointly fits the large and small epidemics, while the three-epidemic design adds a medium epidemic. All designs share $(\mathcal{R}_0, \nu, t_0, c^*)$ across epidemic trajectories. Diagonal correlations are omitted because they are equal to one.

Parameter pair	Two epidemics	Three epidemics
$\mathcal{R}_0-t_0$	-0.68	-0.62
$\mathcal{R}_0-\nu$	-0.04	0.16
$\mathcal{R}_0-c^*$	-0.39	-0.26
$t_0-\nu$	0.39	0.31
t_0-c^*	0.12	0.15
$\nu-c^*$	0.39	0.49

3.6 Chapter Summary, Discussion and Conclusion

3.6.1 Summary of results

The results show that the move from one epidemic to two epidemics leads to a clear improvement in practical identifiability. Profile-likelihood confidence intervals became much narrower for all four parameters, with the largest gains for ν and c^* , the two parameters that were most strongly confounded in Chapter 2. The Hessian condition numbers also improved sharply, showing that the joint likelihood surface is much better conditioned in the two-epidemic setting. In simple terms, the parameter space becomes less flat and less unstable, so estimation is more reliable.

The correlation analysis explains why this happens. In the single-epidemic case, the inverse-Hessian correlation between ν and c^* is very strong (Table 3.3), reflecting the familiar ridge along which heterogeneity and intervention strength can trade off against each other. In the two-epidemic case, this dependence is greatly reduced. The bivariate likelihood surface in Figures 3.3 and 3.4 make the same point visually: the long diagonal ridge seen for one epidemic becomes much tighter and more compact when two epidemics are fitted jointly. This is the clearest

evidence that the extra epidemic provides information in a new direction and helps separate the effects of heterogeneity from those of NPIs.

Appendix C strengthens this point by showing that the same pattern persists across a range of intervention strengths. In particular, Table C.1 shows that the single-epidemic ν - c^* correlation remains very high at all three NPI levels (0.93, 0.96, and 0.90 for $c^* = 0.20$, 0.30, and 0.40, respectively), whereas the two-epidemic correlations are much smaller under strong and moderate control (0.23 and 0.32) but remain higher under weak control (0.68). This shows that joint inference does not remove the ridge equally in every setting: the reduction is strongest when the two trajectories are most informative about depletion differences, and weakest when the intervention is too weak to create as much contrast.

Section 3.5 extended the analysis to 3-epidemics by inserting a “Medium” epidemic (intermediate seed size) alongside the Large and Small. The additional trajectory brought only modest further improvements: CI widths for ν and c^* each shrank by roughly 6-7 % ($\approx 1.1\times$), and those for R_0 and t_0 by 17-22 % (≈ 1.2 - $1.3\times$), relative to the two-epidemic baseline. Condition numbers and the correlation structure were essentially unchanged (Figure 3.7 and Table 3.5).

At the same time, not every difficulty disappears. The negative relationship between $\mathcal{R}_0$ and t_0 remains fairly strong even in the two-epidemic analysis. This suggests that the trade-off between transmission intensity and the timing of onset of behavioural change is a more persistent feature of the model structure. So the main gain from multi-epidemic inference is not that every parameter becomes fully independent, but that the most problematic ridge, the ν - c^* ridge, is substantially weakened.

3.6.2 Discussions and Additional results

The appendices help explain and test the main chapter results from different angles.

First, Appendix C gives two complementary sensitivity checks for this chapter’s result. On one hand, it repeats the profile-likelihood analysis for two additional intervention levels, $c^* \in \{0.20, 0.40\}$ alongside the baseline $c^* = 0.30$, and shows that the qualitative gain from joint fitting persists across all three NPI strengths. Table C.1 is especially useful here because it shows directly how the ν - c^* correlation changes with intervention strength: the reduction from one to two epidemics is largest at $c^* = 0.20$, remains large at $c^* = 0.30$, and is much smaller at $c^* = 0.40$. On the other hand, Appendix C studies sensitivity to initial conditions by varying the initial infected population over $I_i(0) \in \{20, 40, 80, 160, 320, 400\}$. This second check is important for interpreting the baseline Large–Small design: it shows that the correlation structure depends on

how much early-growth and depletion contrast exists between the epidemics. In particular, the baseline choice $l_0 = 400$ versus $l_0 = 40$ is the deliberately well-separated pair, chosen to create a strong enough difference in epidemic stage to weaken the ridge substantially. The appendix does not identify a sharp threshold for how separate two epidemics must be, but it does show that the contrast cannot be too small if the goal is to break the ν - c^* compensation meaningfully.

The fold-improvements are largest at strong control ($c^* = 0.20$: $\approx 23.5\times$ for ν ; $\approx 5.8\times$ for c^*) and smallest at weak control ($c^* = 0.40$: $\approx 6.3\times$ for ν ; $\approx 3.7\times$ for c^*), but substantial in every case. Condition numbers confirm the same story: single-epidemic numbers spread over several orders of magnitude at all three NPI levels, while two-epidemic numbers cluster tightly in the tens (Figure C.1). The three NPI levels $c^* \in \{0.20, 0.30, 0.40\}$ were chosen to represent strong, moderate, and weak sustained intervention under the baseline $\mathcal{R}_0 = 3.0$. The first two produce declining incidence after intervention, while the third slows transmission without fully stopping growth. This gives a simple range of simulation settings for checking that the identifiability results are not tied to one particular intervention strength.

Second, Appendix D tests whether the precision gains survive when the model is made more realistic by relaxing some of the parameter sharing assumptions across the two epidemics, using Bayesian inference. Three scenarios are compared.

1. **Baseline: 4-parameter fully shared model** (R_0, ν, t_0, c^*). The Bayesian replication the main-chapter MLE results.
2. **Scenario A: 5-parameter model with epidemic-specific onset times** $t_0^{(1)}, t_0^{(2)}$.
Allowing the two populations to begin changing their behaviour at different times is a realistic relaxation: different regions are seldom locked down on exactly the same day.
3. **Scenario B: 6-parameter model with epidemic-specific R_0 and t_0** . Allowing each epidemic to have its own reproduction number as well

For the baseline four-parameter model, the answer is yes: the Bayesian analysis supports the main likelihood-based result that joint two-epidemic inference gives tighter uncertainty and weaker ν - c^* dependence than single-epidemic inference. This is important because it shows that the precision gain is not estimation method dependent.

However, the higher-dimensional Bayesian scenarios also show the limit of this gain. When some parameters are allowed to become epidemic-specific, the benefit of joint inference starts to weaken. In Scenario A, where the onset times differ between epidemics, much of the gain is retained, although moderate ν - c^* dependence returns. In Scenario B, where both onset times

and reproduction numbers are epidemic-specific, the ν - c^* ridge becomes strong again. This shows that the success of the multi-epidemic design depends strongly on shared structure across epidemics. If too much flexibility is introduced, the model can again trade off heterogeneity against intervention strength.

3.6.3 Conclusion

The results of this chapter support a simple message: the main problem in Chapter 2 was that the model could not have all parameters identified using a single epidemic time series as it does not contain enough distinct information to separate the effects of susceptibility heterogeneity from the effects of NPIs. By analysing two related epidemics jointly, we change the geometry of the estimation problem. The second epidemic acts like an additional constraint, and this breaks much of the compensation that is present in the single-epidemic case.

Our simulation studies show that joint-epidemic inference is not just a larger sample size argument; it is a change in the kind of information available for estimation. The baseline choice of a Large epidemic ($I_0 = 400$) and a Small epidemic ($I_0 = 40$) makes this concrete: the gain comes from creating a marked contrast in early growth and depletion history while keeping the shared parameters fixed. Appendix C supports this interpretation directly by showing that the correlation structure is sensitive to the initial seed size, even though the dominant single-epidemic ν - c^* confounding remains.

The chapter also shows that there is a balance to be struck between realism and identifiability. A fully shared model may be idealised, but it gives strong gains in precision. Allowing a small amount of epidemic-specific variation may still be acceptable, as seen in Scenario A, Appendix D. But once too many epidemic-specific parameters are introduced, the cross-epidemic information that was helping to identify ν and c^* is partly lost. This is a useful warning for later empirical work: more flexible models are not automatically better if that flexibility destroys the information needed to estimate the quantities of interest.

Another important finding is that adding a third epidemic gives only modest extra benefit. The two-epidemic design already captures most of the available gain. The third epidemic gives some further narrowing of intervals, but the improvements are small and the correlation structure changes little. This suggests that, in the present framework, two well-contrasted epidemics may already be enough to recover most of the identifiability benefit.

Chapter 4

Distribution Discretisation Methods

4.1 Introduction and Motivation

This chapter develops numerical methods for discretising continuous probability distributions of individual susceptibility in the heterogeneous-susceptibility SEIR model introduced in Chapter 1. The continuous susceptibility-indexed system is given in equations (1.64)–(1.67), with force of infection defined in equation (1.68). The aggregate compartment sizes are obtained by integrating the corresponding susceptibility-indexed densities, as defined in equations (1.58)–(1.59). These equations are therefore not repeated here.

Let $q(x)$ denote the initial probability density of susceptibility $x > 0$. As established in Chapter 1, susceptibility is normalised to have mean one, and its coefficient of variation is denoted by ν ; see equations (1.51)–(1.53). The purpose of the present chapter is to replace $q(x)$ by a finite susceptibility grid

$$\{(q_i, x_i)\}_{i=1}^k,$$

where

$$q_i > 0, \quad x_i > 0, \quad \sum_{i=1}^k q_i = 1,$$

and a prescribed coefficient of variation ν :

$$\nu = \sqrt{\sum_{i=1}^k q_i (x_i - 1)^2}, \quad \sum_{i=1}^k q_i x_i = 1.$$

Here q_i is the population proportion assigned to susceptibility group i , and x_i is the representative susceptibility of that group.

Let $S_i(t)$, $E_i(t)$, $I_i(t)$, and $R_i(t)$ denote the numbers of susceptible, exposed, infectious, and removed individuals in susceptibility group i . These are group-specific compartment counts and

therefore retain the subscript i . Their aggregate totals are

$$S(t) = \sum_{i=1}^k S_i(t), \quad E(t) = \sum_{i=1}^k E_i(t), \quad I(t) = \sum_{i=1}^k I_i(t), \quad R(t) = \sum_{i=1}^k R_i(t). \quad (4.1)$$

The force of infection experienced by an individual with unit susceptibility is

$$\lambda_k(t) = \frac{\beta(t)}{N} \{\rho E(t) + I(t)\} = \frac{\beta(t)}{N} \sum_{j=1}^k \{\rho E_j(t) + I_j(t)\}, \quad \beta(t) = \beta c(t), \quad (4.2)$$

where $\rho \in [0, 1]$ is the relative infectiousness of exposed individuals and $c(t)$ is the contact multiplier. In the no-intervention trajectory validation considered later in this chapter, $c(t) \equiv 1$, so that $\beta(t) \equiv \beta$.

The finite susceptibility-stratified SEIR system is

$$\begin{aligned} \frac{dS_i(t)}{dt} &= -x_i \lambda_k(t) S_i(t), \\ \frac{dE_i(t)}{dt} &= x_i \lambda_k(t) S_i(t) - \delta E_i(t), \\ \frac{dI_i(t)}{dt} &= \delta E_i(t) - \gamma I_i(t), \\ \frac{dR_i(t)}{dt} &= \gamma I_i(t), \quad i = 1, \dots, k. \end{aligned} \quad (4.3)$$

When the initially exposed and infectious individuals are distributed proportionally across susceptibility groups, the initial conditions are

$$\begin{aligned} S_i(0) &= S_0 q_i, \\ E_i(0) &= E_0 q_i, \\ I_i(0) &= I_0 q_i, \\ R_i(0) &= 0, \end{aligned} \quad S_0 = N - E_0 - I_0, \quad i = 1, \dots, k. \quad (4.4)$$

These definitions ensure that

$$S(0) = S_0, \quad E(0) = E_0, \quad I(0) = I_0, \quad R(0) = 0,$$

and hence

$$S(0) + E(0) + I(0) + R(0) = N.$$

For a mean-one Gamma susceptibility distribution, the continuous susceptibility-indexed system admits the exact low-dimensional reduction given in equations (1.78)–(1.81), with β

replaced by $\beta(t) = \beta c(t)$ when interventions are included. This reduced Gamma system evolves the aggregate variables S, E, I, R and was used in Chapters 2 and 3. Its availability relies on the shape-preserving property of the Gamma distribution and does not extend directly to an arbitrary susceptibility distribution.

The finite-group system (4.3) and the reduced Gamma system must therefore be distinguished. The former is a numerical approximation to the continuous susceptibility-indexed model and can be used with Gamma, Lognormal, or other positive susceptibility distributions. The latter is an exact four-equation closure available specifically under Gamma-distributed susceptibility.

This distinction motivates the present chapter. Chapter 5 compares Gamma and Lognormal susceptibility distributions with the same mean and coefficient of variation and examines the effects of distributional misspecification on inference and forecasting. Such comparisons cannot be performed using the reduced Gamma system alone, because the full shape of each susceptibility distribution must be retained. The methods developed here construct the finite grids required to simulate both families within the common system (4.3).

The discretisation must satisfy three requirements. First, it should preserve the target mean susceptibility, which is taken to be one so that $\mathcal{R}_0$ retains the same interpretation as the baseline reproduction number. Second, it should preserve the target coefficient of variation ν , since ν is the main parameter used throughout the thesis to quantify susceptibility heterogeneity. Third, it should preserve positivity and ordering of the representative susceptibility values, because negative or non-monotone susceptibility classes would not be epidemiologically meaningful.

This chapter has two primary objectives:

1. To develop a *distribution-agnostic* discretisation framework for positive susceptibility distributions that preserves the target mean μ_* , matches the target coefficient of variation ν where numerically feasible, and maintains positivity and ordering of the representative susceptibility values.
2. To instantiate this framework for the Gamma distribution as the main benchmark, because the discretised Gamma system can be compared against the reduced Gamma model in equations (1.78)–(1.81). The same framework is then used for the Lognormal distribution in Chapter 5, where distributional misspecification is studied explicitly.

The resulting schemes are designed to preserve the key statistical moments used in the thesis, namely a target mean of 1 and a target variance equivalent to the specified coefficient of variation ν . To discretise a positive susceptibility distribution into k groups with representatives x_i and probabilities q_i , we use a three-stage pipeline:

1. **CDF binning:** partition the cumulative distribution into k probability bins to determine the group probabilities $\{q_i\}$. In this chapter, both equal and linear probability spacing are used, with an even number of groups for consistency with the split–multiplicative transformation.
2. **Representative selection:** assign each group a representative susceptibility value x_i , chosen either as the conditional expectation within the bin or as the midpoint on the probability scale.
3. **Moment correction and transformation:** first enforce the target mean $\mu_\star$, then apply a calibration transformation to match the target variance, equivalently the target coefficient of variation ν , while preserving positivity and ordering.

The transformation stage is the main methodological contribution of the chapter. We use two complementary algorithms:

1. **Split–multiplicative (SM):** a two-factor rescaling applied to the lower and upper halves of the grid. The factors are obtained by solving a quadratic equation that enforces the target variance while preserving the target mean. Candidate roots are accepted only if they give positive, ordered, and numerically stable susceptibility representatives.
2. **Log–affine (LA):** a monotone transformation of the form

$$x'_i = \exp(A + B \log x_i),$$

where A enforces the mean constraint and B is chosen by solving a scalar equation for the variance constraint. This transformation gives a smooth proportional adjustment of all susceptibility representatives and preserves positivity by construction.

The chapter is organised as follows. Section 4.2 introduces the distribution-agnostic moment-transformation stage, including the split–multiplicative and log–affine algorithms. Section 4.3 instantiates the framework for the Gamma family. Section 4.4 presents Gamma numerical illustrations for the two calibration methods. Finally, Section 4.5 validates the aggregate trajectories S, E, I, R from the discretised Gamma system against the reduced Gamma system in equations (1.78)–(1.81). This validation is important because it checks that the discretised system reproduces the dynamics of the reduced model before the same discretisation framework is used in Chapter 5 to compare Gamma and Lognormal susceptibility distributions and to study distributional misspecification.

4.2 Distribution-Agnostic Moment Algorithms

Let X be a positive continuous random variable with cumulative distribution function F . We propose a distribution-agnostic method for discretising X using a positive finite grid

$$\{(q_i, x_i)\}_{i=1}^k,$$

where

$$q_i > 0, \quad x_i > 0, \quad \sum_{i=1}^k q_i = 1.$$

The probability masses q_i define the bins of the discretisation, while the representatives x_i are susceptibility values assigned to those bins. The aim is to construct a grid whose empirical mean is a prescribed target $\mu_\star$,

$$\sum_{i=1}^k q_i x_i = \mu_\star,$$

and whose empirical coefficient of variation is a prescribed value $\nu > 0$. Equivalently, the target variance is

$$\sigma_\star^2 = \nu^2 \mu_\star^2.$$

After the probability masses have been chosen, the calibration stage adjusts only the representatives $\{x_i\}$, producing transformed representatives $\{x'_i\}$ satisfying

$$\sum_{i=1}^k q_i (x'_i - \mu_\star)^2 = \sigma_\star^2,$$

while preserving positivity and ordering of the susceptibility grid.

Throughout this chapter, k is taken to be even. This convention is used for both equal and linear probability spacing and is also required by the split-multiplicative transformation in Section 4.2.1, which divides the grid into lower and upper halves.

The discretisation proceeds in three stages.

1. **CDF binning.** Partition the cumulative distribution into k contiguous bins with probability masses $\{q_i\}_{i=1}^k$, where

$$q_i > 0, \quad \sum_{i=1}^k q_i = 1.$$

Two spacing schemes are considered throughout.

- *Equal spacing.* All bins receive the same probability mass:

$$q_i = \frac{1}{k}, \quad i = 1, \dots, k. \quad (4.5)$$

- *Linear spacing.* Linear spacing assigns smaller probability masses to the extreme bins and larger probability masses to bins near the centre of the distribution. For even k , set $m = k/2$ and define the raw weights

$$w_i = \min(i, k + 1 - i), \quad i = 1, \dots, k.$$

Since

$$\sum_{i=1}^k w_i = m(m + 1),$$

the normalised probabilities are

$$q_i = \frac{\min(i, k + 1 - i)}{m(m + 1)}, \quad i = 1, \dots, k. \quad (4.6)$$

The cumulative bin boundaries are

$$z_0 = 0, \quad z_i = \sum_{j=1}^i q_j, \quad i = 1, \dots, k.$$

The corresponding distribution-specific bin edges are then obtained from the quantile function:

$$a_i = F^{-1}(z_{i-1}), \quad b_i = F^{-1}(z_i), \quad i = 1, \dots, k.$$

2. **Representative selection.** For each bin $[a_i, b_i]$, choose a representative susceptibility value x_i . In this thesis, two choices are used. The first is the conditional expectation within the bin,

$$x_i = \mathbb{E}\{X \mid a_i < X \leq b_i\},$$

and the second is the probability-scale midpoint,

$$x_i = F^{-1}\left(\frac{z_{i-1} + z_i}{2}\right).$$

The conditional expectation is used as the default representative because it uses the average value of X within each probability bin.

3. Moment correction and calibration. The discrete approximation initially constructed from $\{(q_i, x_i)\}$ typically satisfies

$$\sum_{i=1}^k q_i = 1,$$

but may not reproduce the theoretical mean exactly. To enforce the mean constraint

$$\sum_{i=1}^k q_i x_i = \mu_\star,$$

the representative of the last bin is adjusted:

$$x_k \leftarrow x_k - \frac{\sum_{i=1}^k q_i x_i - \mu_\star}{q_k}.$$

Equivalently,

$$x_k \leftarrow \frac{\mu_\star - \sum_{i=1}^{k-1} q_i x_i}{q_k}.$$

This correction minimally perturbs the discretisation while guaranteeing the target mean.

A subsequent calibration transformation then adjusts the representatives to match the target variance

$$\sigma_\star^2 = \nu^2 \mu_\star^2.$$

We use two order-preserving calibration transformations: a *split-multiplicative* rescaling and a *log-affine* map. These are presented in the following subsections.

4.2.1 Split-multiplicative (SM) transformation

The split-multiplicative (SM) transformation is used to adjust a mean-corrected discrete susceptibility grid so that it also matches the target variance. Let

$$\{(q_i, x_i)\}_{i=1}^k$$

denote the discretised distribution after CDF binning, representative selection, and mean correction, so that

$$q_i > 0, \quad \sum_{i=1}^k q_i = 1, \quad x_i > 0, \quad \sum_{i=1}^k q_i x_i = \mu_\star.$$

In this thesis, we use an even number of groups k for both equal and linear probability spacing. This keeps the discretisation conventions consistent across the chapter and allows the SM

transformation to split the grid into two halves:

$$L = \{1, \dots, k/2\}, \quad R = \{k/2 + 1, \dots, k\}.$$

For the two halves, define the partial first and second raw moments

$$X_L = \sum_{i \in L} q_i x_i, \quad X_R = \sum_{i \in R} q_i x_i,$$

and

$$M_L = \sum_{i \in L} q_i x_i^2, \quad M_R = \sum_{i \in R} q_i x_i^2.$$

Since the grid has already been corrected to have mean $\mu_\star$,

$$X_L + X_R = \mu_\star.$$

The SM transformation applies one multiplier to the lower half and another multiplier to the upper half:

$$x'_i = \begin{cases} rx_i, & i \in L, \\ ux_i, & i \in R, \end{cases}$$

where $r > 0$ and $u > 0$. The mean constraint is

$$rX_L + uX_R = \mu_\star.$$

Thus, once r is chosen, the upper-half multiplier is fixed:

$$u(r) = \frac{\mu_\star - rX_L}{X_R}.$$

The target variance is

$$\sigma_\star^2 = \nu^2 \mu_\star^2.$$

Because the SM transformation is constructed to preserve the mean, matching the target variance is equivalent to matching the second raw moment:

$$\sum_{i=1}^k q_i (x'_i)^2 = \mu_\star^2 + \sigma_\star^2.$$

Using the two-half form of the transformation gives

$$r^2 M_L + u(r)^2 M_R = \mu_\star^2 + \sigma_\star^2.$$

Substituting $u(r) = (\mu_\star - r X_L)/X_R$ gives a quadratic equation for the lower-half multiplier r :

$$ar^2 + br + c = 0,$$

where

$$a = M_L + M_R \frac{X_L^2}{X_R^2}, \quad b = -2M_R \mu_\star \frac{X_L}{X_R^2}, \quad c = M_R \frac{\mu_\star^2}{X_R^2} - (\mu_\star^2 + \sigma_\star^2).$$

Equivalently, defining $T_\star = \mu_\star^2 + \sigma_\star^2$ gives $c = M_R \mu_\star^2 / X_R^2 - T_\star$.

The discriminant is $D = b^2 - 4ac$. If $D \leq 0$, or if the quadratic coefficients are not finite, the SM correction is not forced. Instead, identity multipliers are used:

$$r = u = 1.$$

In this case, the method returns the mean-corrected grid. This avoids replacing a stable discretisation by a numerically invalid correction.

When $D > 0$, the two candidate roots are

$$r_1 = \frac{-b + \sqrt{D}}{2a}, \quad r_2 = \frac{-b - \sqrt{D}}{2a}.$$

Both roots are evaluated, and a root is accepted only if it satisfies explicit admissibility conditions.

For a candidate r , the corresponding upper-half multiplier is

$$u(r) = \frac{\mu_\star - r X_L}{X_R}.$$

Remark. The existence of real roots is not sufficient for a valid susceptibility discretisation. The purpose of the admissibility conditions is to ensure that the moment correction remains epidemiologically and numerically meaningful. A candidate root is admissible only if:

1. both r and $u(r)$ are finite and positive;
2. both multipliers are not excessively large or small, expressed as

$$|\log r| \leq C, \quad |\log u(r)| \leq C,$$

where $C > 0$ is a fixed numerical cap;

3. the corrected grid remains ordered across the split:

$$\max_{i \in L} \{rx_i\} \leq \min_{i \in R} \{u(r)x_i\}.$$

The logarithmic cap penalises inflation and deflation symmetrically. For example, r and $1/r$ have the same absolute log-distance from one. The ordering condition is also sufficient to preserve monotonicity of the full grid. Indeed, multiplication by a positive constant preserves ordering within each half. Therefore, any possible loss of ordering can only occur at the boundary between the lower and upper halves.

If both candidate roots are admissible, the method selects the least-distorting root, measured by

$$S(r, u) = |\log r| + |\log u|.$$

This selects the admissible correction closest to the identity transformation, while still matching the target moments. If no admissible root exists, the split-multiplicative correction is not forced, and the mean-corrected grid is retained. This prevents the transformation from introducing non-positive, non-monotone, or numerically unstable representatives.

After the selected transformation has been applied, a final normalisation may be used to remove negligible numerical drift in the mean:

$$x'_i \leftarrow x'_i \frac{\mu_\star}{\sum_{j=1}^k q_j x'_j}.$$

4.2.2 Log-Affine (LA) transformation: algebra, existence/uniqueness, implementation

Set $y_i = \log x_i$ and define the log-sum-exp moments

$$M(t) = \sum_{i=1}^k q_i e^{ty_i} = \sum_{i=1}^k q_i x_i^t, \quad h(t) = \log M(t).$$

Apply the log-affine map

$$x'_i = e^A x_i^B \iff y'_i = A + B y_i \iff y'_i = \log x'_i,$$

with (A, B) chosen to satisfy

$$\sum_i q_i x'_i = \mu_\star, \quad \sum_i q_i (x'_i - \mu_\star)^2 = \sigma_\star^2.$$

Mean constraint.

$$\sum_i q_i x'_i = e^A M(B) = \mu_\star \implies A(B) = \log \mu_\star - \log M(B) = \log \mu_\star - h(B).$$

Variance constraint.

$$\mathbb{E}[(X')^2] = e^{2A} M(2B) = \mu_\star^2 \frac{M(2B)}{M(B)^2}.$$

Since $\text{Var}(X') = \mathbb{E}[(X')^2] - \mu_\star^2$, matching $\sigma_\star^2$ is equivalent to

$$\frac{M(2B)}{M(B)^2} = 1 + \frac{\sigma_\star^2}{\mu_\star^2} = 1 + \nu^2 \iff g(B) := h(2B) - 2h(B) - \log(1 + \nu^2) = 0.$$

Existence and uniqueness.

Let

$$h(t) = \log \left(\sum_{i=1}^k q_i e^{t y_i} \right), \quad y_i = \log x_i,$$

and define

$$g(B) = h(2B) - 2h(B) - \log(1 + \nu^2), \quad B \geq 0.$$

At $B = 0$,

$$g(0) = h(0) - 2h(0) - \log(1 + \nu^2) = -\log(1 + \nu^2) < 0,$$

because $\sum_i q_i = 1$, so $h(0) = 0$.

To study monotonicity, differentiate g :

$$g'(B) = 2h'(2B) - 2h'(B).$$

Since h is a log-sum-exp function, it is convex. Hence h' is non-decreasing, and therefore

$$h'(2B) \geq h'(B), \quad B \geq 0.$$

It follows that

$$g'(B) \geq 0.$$

Thus g is non-decreasing on $[0, \infty)$, and it is strictly increasing unless the grid is degenerate, meaning all y_i are equal.

It remains to determine whether $g(B)$ crosses zero. Define

$$y_{\max} = \max_{1 \leq i \leq k} y_i, \quad I_{\max} = \{i : y_i = y_{\max}\}, \quad Q_{\max} = \sum_{i \in I_{\max}} q_i.$$

As $B \rightarrow \infty$, the largest value of y_i dominates the log-sum-exp expression. Therefore,

$$h(B) = \log \left(\sum_{i=1}^k q_i e^{By_i} \right) = By_{\max} + \log Q_{\max} + o(1),$$

and similarly

$$h(2B) = 2By_{\max} + \log Q_{\max} + o(1).$$

Hence

$$\begin{aligned} \lim_{B \rightarrow \infty} g(B) &= \lim_{B \rightarrow \infty} \left\{ h(2B) - 2h(B) - \log(1 + \nu^2) \right\} \\ &= (2By_{\max} + \log Q_{\max}) - 2(By_{\max} + \log Q_{\max}) - \log(1 + \nu^2) \\ &= -\log Q_{\max} - \log(1 + \nu^2) \\ &= \log \left(\frac{1}{Q_{\max}} \right) - \log(1 + \nu^2). \end{aligned}$$

Consequently,

$$\lim_{B \rightarrow \infty} g(B) > 0 \iff 1 + \nu^2 < \frac{1}{Q_{\max}}.$$

Under this condition, $g(0) < 0$, $g(B)$ is continuous and increasing, and $g(B)$ becomes positive for sufficiently large B . Therefore, by the Intermediate Value Theorem, there exists a finite root $B^* > 0$ such that

$$g(B^*) = 0.$$

Because g is strictly increasing for a non-degenerate grid, this root is unique.

If

$$1 + \nu^2 = \frac{1}{Q_{\max}},$$

then the root is approached only in the limit $B \rightarrow \infty$. If

$$1 + \nu^2 > \frac{1}{Q_{\max}},$$

then no finite solution exists. Thus, in practical computation, a finite log-affine solution exists

and is unique whenever

$$1 + \nu^2 < \frac{1}{Q_{\max}}.$$

For equal-probability spacing with a unique maximum representative, $Q_{\max} = 1/k$, so the condition becomes

$$1 + \nu^2 < k.$$

Thus, for a moderate number of groups, the log-affine transformation can match a wide range of target coefficients of variation while preserving positivity and monotonicity.

Remark on numerical bracketing. The scalar equation

$$g(B) = h(2B) - 2h(B) - \log(1 + \nu^2) = 0$$

is solved only after a valid bracket has been found. Since

$$g(0) = -\log(1 + \nu^2) < 0,$$

the lower endpoint can be fixed at $B_{\text{lo}} = 0$. The numerical task is therefore to find an upper endpoint $B_{\text{hi}} > 0$ such that

$$g(B_{\text{hi}}) > 0.$$

Starting from a moderate value of B_{hi} , the upper endpoint is increased geometrically until a sign change is obtained:

$$g(B_{\text{lo}}) < 0 < g(B_{\text{hi}}).$$

Once this bracket is available, a one-dimensional root-finding method can be used to obtain the unique solution B^* .

This bracketing step is important because the existence proof gives a condition under which a finite root exists, but finite-precision computation may still fail to locate a reliable interval if the requested variance is too large for the chosen grid or if the root is effectively approached only at very large values of B . In such cases, the log-affine correction is not forced. Instead, the method retains the positive mean-corrected grid.

Implementation.

In practice, the log–affine transformation is reduced to solving the scalar equation

$$g(B) = h(2B) - 2h(B) - \log(1 + \nu^2) = 0.$$

Since

$$g(0) = -\log(1 + \nu^2) < 0,$$

we set $B_{\text{lo}} = 0$ and search for $B_{\text{hi}} > 0$ such that

$$g(B_{\text{hi}}) > 0.$$

The upper endpoint is increased geometrically until the sign change

$$g(B_{\text{lo}}) < 0 < g(B_{\text{hi}})$$

is obtained. Once this bracket is found, a one-dimensional root-finding method is applied on $[B_{\text{lo}}, B_{\text{hi}}]$ to obtain B^* . The intercept is then determined from the mean constraint:

$$A^* = \log \mu_\star - h(B^*),$$

and the calibrated representatives are

$$x'_i = \exp(A^* + B^* \log x_i).$$

Remark If no upper endpoint with $g(B_{\text{hi}}) > 0$ is found within a reasonable search range, the variance correction is not forced. This can occur when the requested coefficient of variation is too large for the finite grid, or when the root is approached only at an impractically large value of B . In this case, the method returns the positive mean-corrected grid rather than applying an unstable transformation. The purpose of this fallback is the same as in the split–multiplicative method: the discretisation should preserve positivity, ordering, and numerical stability, even if exact variance matching is not available for the chosen grid.

4.3 Gamma Discretisation: Distribution-specific Ingredients

Let

$$X \sim \text{Gamma}(\alpha, \alpha),$$

where the first $\alpha > 0$ is the shape parameter and the second is the rate parameter. Thus, throughout this chapter, the Gamma distribution is written directly in its mean-one shape–rate parameterisation. Its density is

$$f_{\alpha, \alpha}(x) = \frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x}, \quad x > 0,$$

and its cumulative distribution function is denoted by $F_{\alpha, \alpha}$.

Parameterisation. For $X \sim \text{Gamma}(\alpha, \alpha)$,

$$\mu_\star = \mathbb{E}[X] = \frac{\alpha}{\alpha} = 1, \quad \text{Var}(X) = \frac{\alpha}{\alpha^2} = \frac{1}{\alpha}, \quad \nu^2 = \frac{1}{\alpha}.$$

The discretisation steps is the distribution-agnostic procedure developed in Section 4.2; the remainder of this section records the Gamma-specific ingredients required to instantiate it.

CDF binning (Gamma quantiles). Choose bin probabilities $\{q_i\}_{i=1}^k$ using the same equal or linear CDF binning rules as in Section 4.2, and form cumulative masses $c_i = \sum_{j \leq i} q_j$. The bin edges are Gamma quantiles

$$t_i = F_{\alpha, \alpha}^{-1}(c_i), \quad a_i = t_{i-1}, \quad b_i = t_i,$$

with $t_0 = 0$ and $t_k = +\infty$.

Representatives (bin conditional means). For each bin $[a_i, b_i)$ we use the conditional expectation

$$x_i = \mathbb{E}[X \mid a_i \leq X < b_i] = \frac{\int_{a_i}^{b_i} x f_{\alpha, \alpha}(x) dx}{F_{\alpha, \alpha}(b_i) - F_{\alpha, \alpha}(a_i)}.$$

Using

$$x f_{\alpha, \alpha}(x) = f_{\alpha+1, \alpha}(x),$$

where $f_{\alpha+1,\alpha}$ is the Gamma density with shape $\alpha + 1$ and rate α , gives the closed form

$$x_i = \frac{F_{\alpha+1,\alpha}(b_i) - F_{\alpha+1,\alpha}(a_i)}{F_{\alpha,\alpha}(b_i) - F_{\alpha,\alpha}(a_i)}.$$

Mean correction method Although the construction targets $\mu_\star = 1$, finite precision can produce a small mean drift. We enforce $\sum_i q_i x_i = \mu_\star$ by a minimal adjustment of the final bin:

$$x_k \leftarrow x_k - \frac{\sum_i q_i x_i - \mu_\star}{q_k}.$$

Moment transformation (LA and SM). Finally, we apply the generic transformation step of Section 4.2 with $(\mu_\star, \nu^2) = (1, 1/\alpha)$:

- **LA:** apply $x'_i = \exp(A + B \log x_i)$ with $A(B) = -h(B)$ and solve $g(B) = h(2B) - 2h(B) - \log(1 + \nu^2) = 0$ using $\nu^2 = 1/\alpha$. If the root cannot be bracketed, the implementation returns the mean-normalised initial discretisation without further adjustment.
- **SM:** apply a two-factor half-rescaling (Section 4.2.1) and select an admissible positive solution. If no admissible real solution exists, the code uses identity multipliers and returns the mean-normalised initial discretisation.

Section 4.4 illustrates these constructions numerically, and Section 4.5 validates them by trajectory matching against the Gamma-reduced system in equations (1.78)–(1.81).

4.4 Numerical illustration for the Gamma discretisation

We demonstrate the methods described with some illustrative examples. We take a fairly moderate coefficient of variation (any value of $\text{CV} > 1$ is deemed high). Let us consider $\text{Gamma}(\alpha, \alpha)$ with $\alpha = 0.75$ (mean = 1, $\text{CV} \approx 1.155$) and $k = 10$ groups. We use the two algorithms from Section 4.2: split-multiplicative (SM) and log-affine (LA).

Log-affine (LA) transformation

For LA transformation, the transformation solves

$$g(B) = h(2B) - 2h(B) - \log(1 + \nu^2) = 0,$$

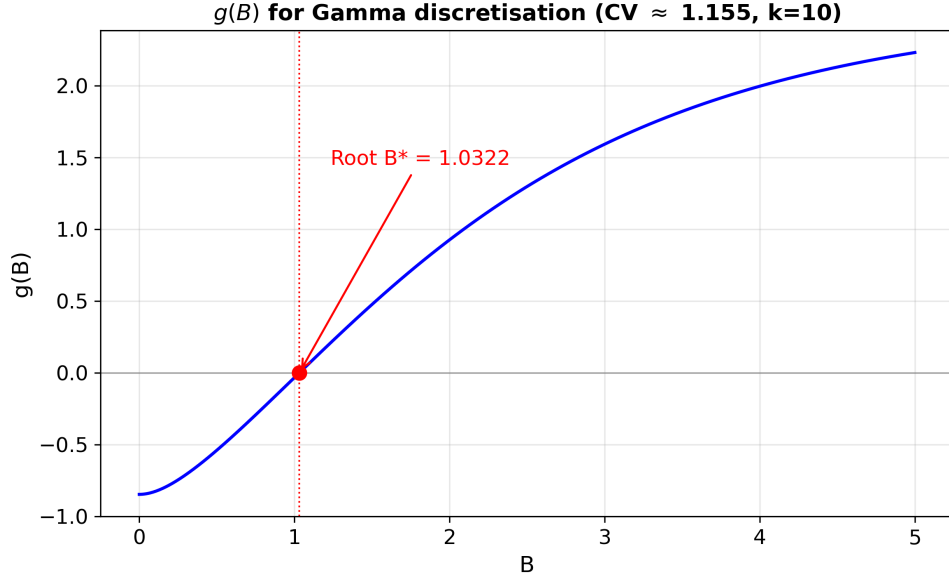


Figure 4.1: $g(B)$ for the discretised Gamma distribution ($CV \approx 1.155$) using the LA transformation of Section 4.2.2. The function starts negative at $B = 0$ and increases monotonically, crossing zero at $B^* \approx 1.0322$.

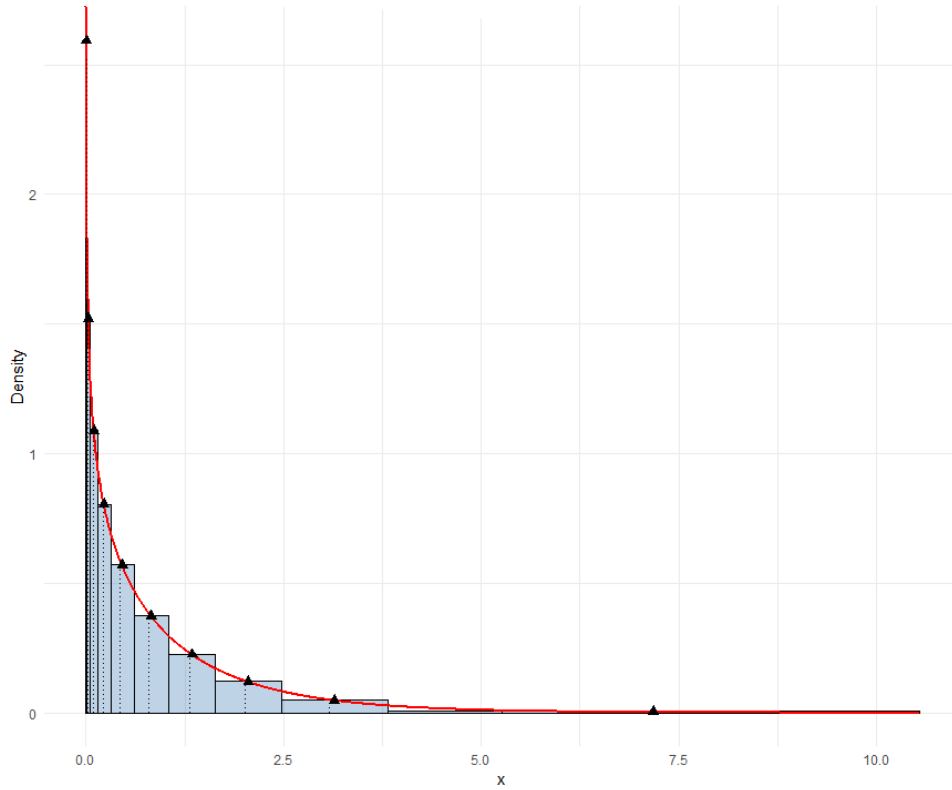


Figure 4.2: Discretised Gamma approximation vs. the true PDF after LA transformation. Blue rectangles represent quantile bins; the red curve is the true Gamma PDF; dashed stems indicate the calibrated representatives x'_i .

yielding a unique root B^* and corresponding $A^* = \log \mu_* - h(B^*)$ with $\mu_* = 1$. The calibrated representatives are

$$x'_i = \exp(A^* + B^* \log x_i), \quad x'_i \leftarrow x'_i / \sum_j q_j x'_j.$$

Figure 4.1 shows the monotone $g(B)$ and its unique root. Table 4.1 summarizes the transformation parameters for the two CDF binning options at $\text{CV} \approx 1.155$ and $k = 10$. Figure 4.2 gives the quantile–rectangle fitted against the true PDF. In all cases, LA preserves the shape and enforces the target mean and variance exactly.

Table 4.1: Log–affine transformation parameters and empirical checks ($\text{CV} \approx 1.155$, $k = 10$).

Spacing	A^*	B^*	Empirical Mean	Empirical CV
Linear	−0.0172	1.0323	1.0000	1.1549
Equal	−0.0451	1.0834	1.0000	1.1549

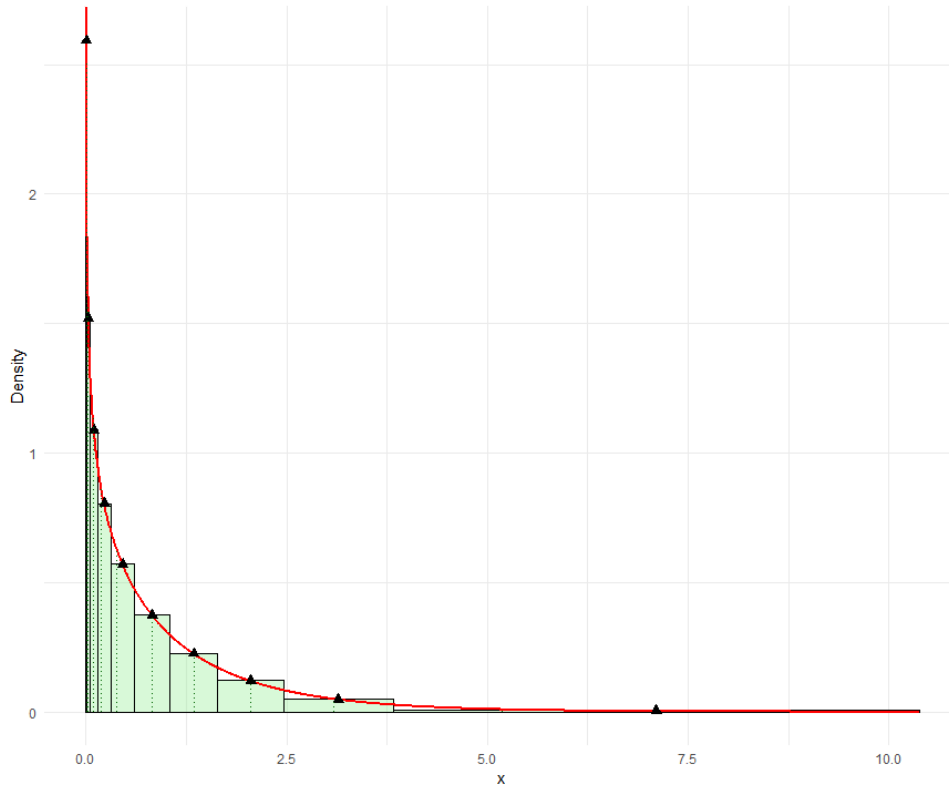


Figure 4.3: Gamma PDF fit after SM transformation. Light-green rectangles show quantile bins; the red curve is the true Gamma PDF; dotted stems mark the calibrated x'_i . The transformation used $r \approx 1.1749$ and $u \approx 1.0199$.

Split–multiplicative (SM) transformation

After constructing bin representatives via conditional expectations and applying the mean correction, we solve the SM quadratic (Section 4.2.1) and obtain the two scaling factors (r, u)

across two partitions, iteratively refined until convergence:

$$x'_i \leftarrow \begin{cases} r \cdot x_i, & i \leq k/2, \\ u \cdot x_i, & i > k/2. \end{cases}$$

Finally, normalization ensures $\sum_j q_j x'_j = 1$.

Table 4.2 reports the transformation parameters for to CDF binning options at $CV \approx 1.155$ and $k = 10$. As with LA, SM preserves monotonicity and matches the target moments exactly.

Table 4.2: Split-multiplicative transformation parameters and empirical checks ($CV \approx 1.155$, $k = 10$).

Spacing	r	u	Empirical Mean	Empirical CV
Linear	1.1751	1.0199	1.0000	1.1549
Equal	1.4941	1.0441	1.0000	1.1549

Visual comparison of calibrated representatives

Figures 4.4 and 4.5 compare the original conditional-mean representatives with the calibrated grids produced by the LA and SM transformations under equal and linear spacing, respectively. In both cases, calibration leaves the representatives close to their original values and preserves the increasing order across groups. This shows that moment matching is achieved without materially distorting the discrete susceptibility grid.

The equal-spacing grid shows somewhat larger adjustments in the upper groups, because the final bins contain more extreme representative values and therefore play a stronger role in matching the target variance. The linear-spacing grid requires less visible correction, with the calibrated and original curves almost overlapping throughout. In both settings, LA and SM give very similar results, although LA produces a smoother adjustment across the whole grid, while SM acts through separate rescalings of the lower and upper halves.

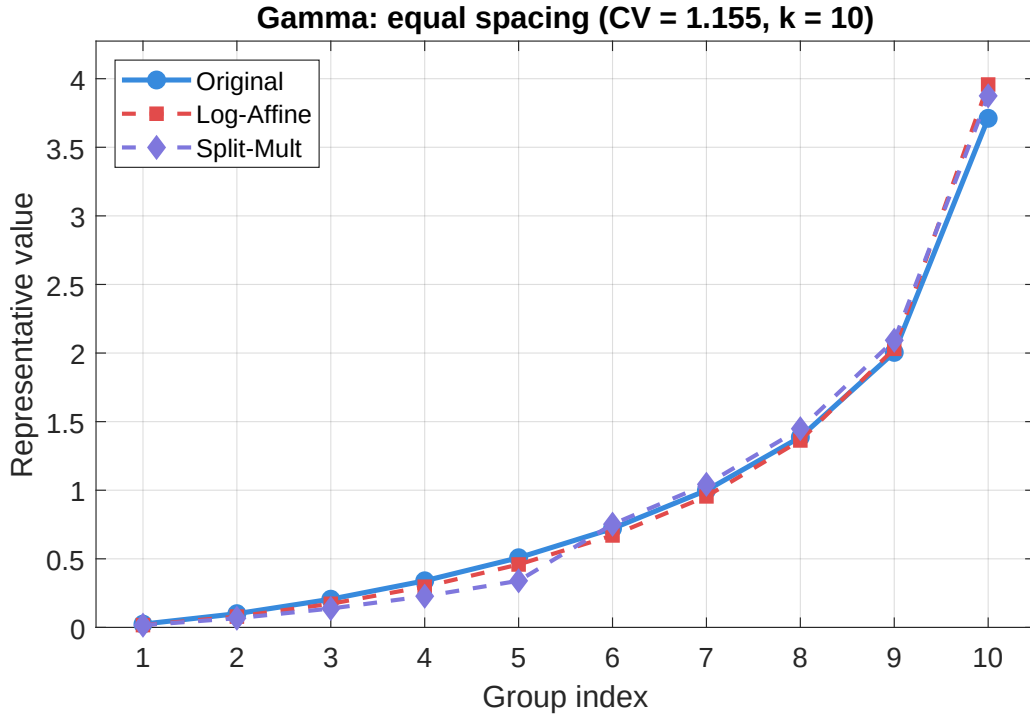


Figure 4.4: Gamma discretisation with equal probability spacing ($CV \approx 1.155$, $k = 10$). The blue solid curve shows the original conditional-mean representatives, while the red dashed and purple dashed curves show the calibrated representatives obtained using the log-affine (LA) and split-multiplicative (SM) transformations, respectively. Both transformations preserve monotonicity and remain close to the original grid, with the main adjustments occurring in the upper susceptibility groups in order to match the target mean and coefficient of variation exactly.

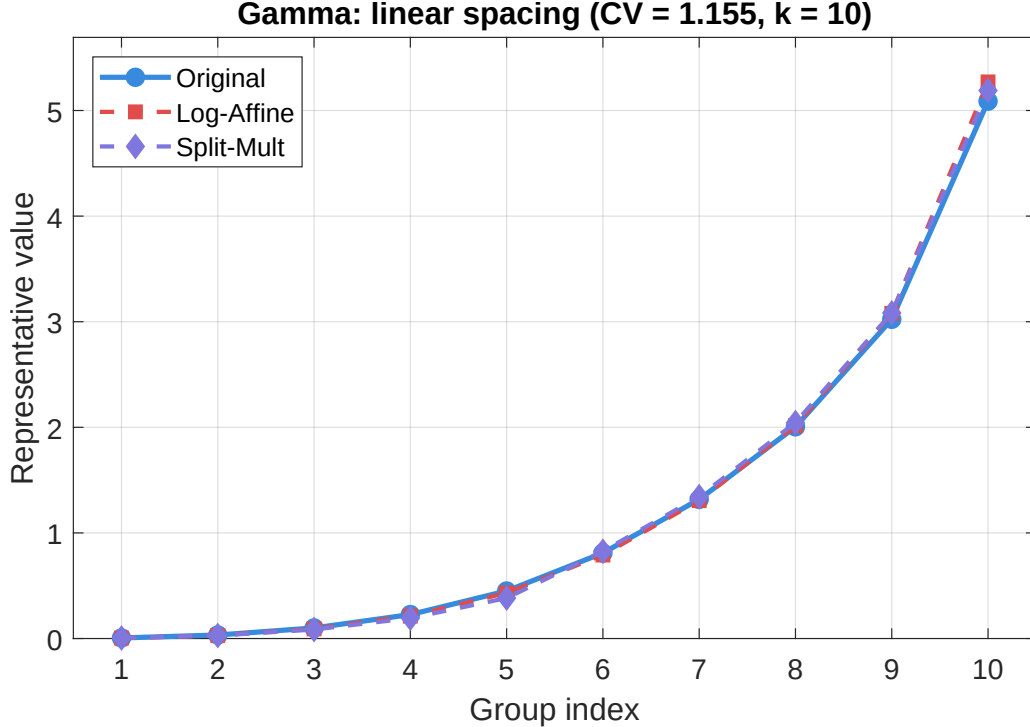


Figure 4.5: Gamma discretisation with linear probability spacing ($CV \approx 1.155$, $k = 10$). The blue solid curve shows the original conditional-mean representatives, while the red dashed and purple dashed curves show the calibrated representatives from the log-affine (LA) and split-multiplicative (SM) transformations, respectively. The three curves are nearly indistinguishable over most of the grid, showing that both calibration methods preserve the shape and ordering of the initial discretisation while enforcing the target moments.

4.5 Trajectory Matching

To evaluate the accuracy of the proposed discretisation schemes, we test whether the finite-group Gamma model (4.3) reproduces the epidemic trajectories generated by the exact reduced Gamma system in equations (1.78)–(1.81). The comparison therefore assesses whether the moment-matched finite susceptibility grid provides an accurate numerical approximation to the continuous Gamma susceptibility model.

Distribution-level diagnostics are reported in Appendix E, Section E.5. Those checks assess how accurately the discrete probability grid represents the target distribution, whereas the analysis here assesses agreement at the level of epidemic trajectories.

Objective

The objective is to verify that the moment-preserving discretisation, when used in system (4.3), produces aggregate trajectories that agree with those of the exact reduced Gamma model (1.78)–(1.81).

Model Setup

Both the discretised heterogeneous SEIR model and the reduced model are initialized with identical SEIR model parameters:

- total population size N and basic reproduction number

$$R_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right), \quad (4.7)$$

where ρ is the relative infectiousness of exposed individuals, δ is the progression rate from the exposed to the infectious compartment, γ is the removal rate from the infectious compartment, and β is the baseline transmission rate. The parameter values are those used in Chapter 2, Table 2.1.

- initial conditions (E_0, I_0) distributed among k susceptibility classes proportionally to q_i .

The heterogeneous system integrates $4k$ ordinary differential equations (for S_i, E_i, I_i, R_i), while the reduced system integrates the four aggregated equations using the same parameters and the theoretical variance $\nu^2 = 1/\alpha$.

Comparison Metrics

Outputs from both systems—total $S(t)$, $E(t)$, $I(t)$, and $R(t)$, are compared using:

1. **Absolute and mean absolute differences:**

$$\max_t |I_{\text{disc}}(t) - I_{\text{red}}(t)|, \quad \frac{1}{T} \int_0^T |I_{\text{disc}}(t) - I_{\text{red}}(t)| dt.$$

2. **Squared deviation metrics:** quantifying the overall trajectory discrepancy via integrated squared error.

3. **Peak-based metrics:**

$$\Delta t_{\text{peak}} = t_{\text{disc}}^{(\max)} - t_{\text{red}}^{(\max)}, \quad \Delta I_{\text{peak}} = I_{\text{disc}}^{(\max)} - I_{\text{red}}^{(\max)}.$$

Graphical comparisons are also provided by overlaying the $S(t)$, $E(t)$, $I(t)$, and $R(t)$ trajectories from the discretised and reduced models, to assess the agreement visually.

Simulation setup.

We run $k \in \{10, 20, 30, 40, 50, 100\}$ groups over four heterogeneity levels $\text{CV} \in \{0.5, 1, 1.5, 2\}$, using identical epidemiological parameters and initial conditions in both models. All results below use the 'linear' spacing and conditional expectation representatives described in 4.3 and apply both log-affine or split-multiplicative transformations.

Results

Table 4.3: Trajectory discrepancies as a function of heterogeneity ($k = 20$ groups). For each coefficient of variation ν , we report $\max_t |I_{\text{disc}}(t) - I_{\text{red}}(t)|$ and the peak timing difference Δt_{peak} (in days) between the discretised and reduced models, for both transformation methods.

ν	log_affine		split_mult	
	max $ I $	diff Δt_{peak}	max $ I $	diff Δt_{peak}
0.500	8	0	10	0
1.000	33	0	63	0
1.500	58	0	144	0
2.000	73	1	205	1

Max infected cases and Peak timing with k

For almost all scenarios considered, the discretised model and the reduced model peaked on the exact same day and exhibited a very similar maximum number of infected cases, with minor deviations appearing only at the coarser resolutions of $k = 10$ and $k = 20$. For instance, Tables

Table 4.4: Trajectory discrepancies as a function of heterogeneity ($k = 10$ groups). For each coefficient of variation ν , we report $\max_t |I_{\text{disc}}(t) - I_{\text{red}}(t)|$ and the peak timing difference Δt_{peak} (in days) between the discretised and reduced models. Missing values (—) for the split-multiplicative (SM) method at $\nu = 2.0$ (Gamma shape $\alpha = 0.25$) occur because the transformation quadratic fails to yield an admissible real root in this extreme high-variance regime at coarse resolutions.

ν	log_affine		split_mult	
	max $ I $	diff Δt_{peak}	max $ I $	diff Δt_{peak}
0.500	20	0	27	0
1.000	102	0	195	0
1.500	186	0	444	1
2.000	244	1	—	—

Table 4.5: Trajectory discrepancies for $k \in \{40, 50, 100\}$ and $\nu \in \{1.5, 2.0\}$. For each pair (ν, K) , we report $\max_t |I_{\text{disc}}(t) - I_{\text{red}}(t)|$ and the peak timing difference Δt_{peak} (days) for both transformation methods.

ν	k	log_affine		split_mult	
		max $ I $	diff Δt_{peak}	max $ I $	diff Δt_{peak}
1.5	40	23	0	48	0
1.5	50	24	0	35	0
1.5	100	27	0	26	0
2.0	40	25	0	68	1
2.0	50	27	0	48	1
2.0	100	30	0	29	0

4.3 and 4.4 show that at $k = 20$, trajectory gaps are small with peak-timing differences of 0–1 day; the LA transformation is uniformly tighter than SM (e.g., at $\nu = 2.0$: 73 vs 205 cases). At $k = 10$, the gaps are predictably larger, yet timing discrepancies remain restricted to ≤ 1 day. As highlighted above, the SM method completely fails to find an admissible solution at the high-variance threshold ($\nu = 2.0$, $\alpha = 0.25$) for $k = 10$. Thus, errors decrease rapidly with k for lower or moderate CVs, with LA consistently proving to be at least as accurate and significantly more numerically robust. This stable convergence is further evidenced in Table 4.5 for finer grids ($k = 40, 50, 100$).

Error reduction relative to $k = 10$. Figure 4.6 recasts the same quantity as percent reduction against the $k = 10$ baseline. Increasing to $k = 20$ removes the bulk of trajectory error for all CVs; additional groups yield progressively smaller gains. **LA** and **SM** show practically indistinguishable improvement rates across CVs, with the same SM omission at CV = 2, $k = 10$.

Trajectory at $k = 20$. To visualize the remaining discrepancies, Figure 4.7 overlays the discretised and reduced infection curves for all four CVs at $k = 20$ while Figure 4.8 shows the

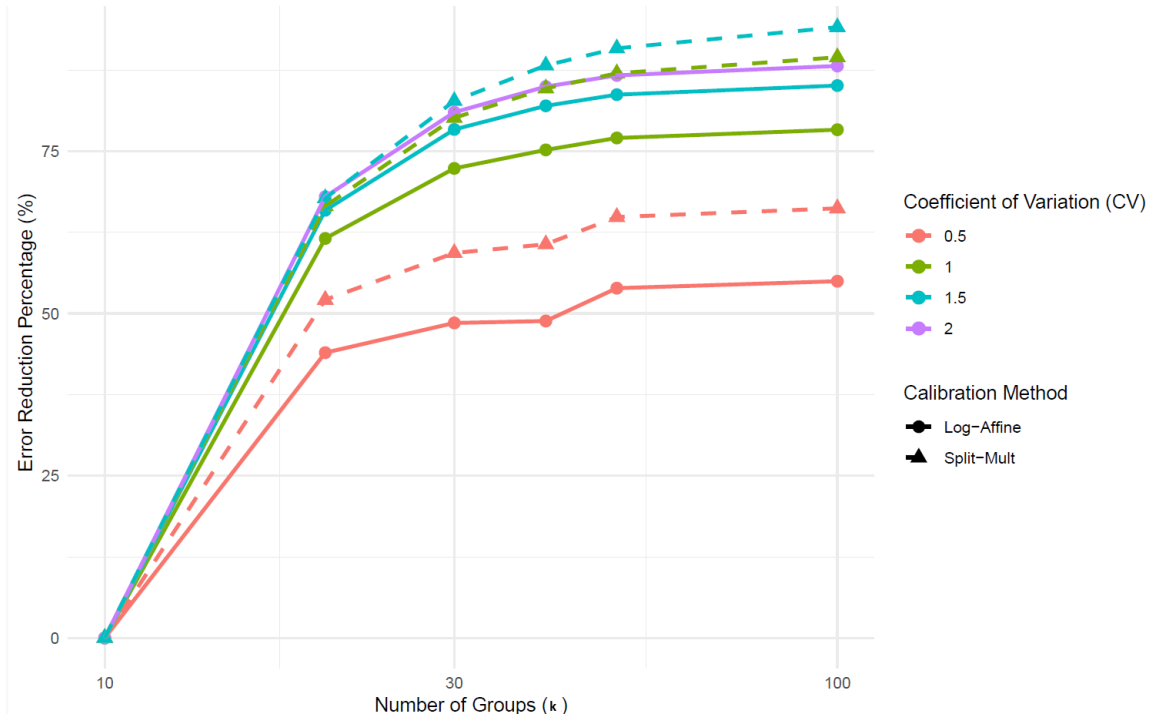


Figure 4.6: Percentage reduction in mean absolute trajectory error relative to $k = 10$, by CV and transformation. Moving from $k = 10$ to $k = 20$ yields the largest gain; thereafter, improvements taper. The SM curve for $CV = 2$ is omitted because the algorithm failed to provide a solution at $k = 10$.

S, E, I, R dynamics also for $k = 20$ for LA with $CV=1.5$. Across both transformations, curves are nearly indistinguishable: peaks align closely in timing and magnitude, and the full outbreak shapes coincide to within plotting resolution.

Summary. (i) Trajectory errors decay rapidly with k and exhibit clear diminishing returns after ~ 20 groups; (ii) log-affine and split-multiplicative transformations perform essentially the same wherever the SM solution exists; (iii) at $CV = 2$, $k = 10$ the SM quadratic has no real root, while log-affine remains robust; and (iv) at $k = 20$ the discretised dynamics are already in very close agreement with the analytical reduced model.

Practical implication. A discretisation with $k \approx 20$ is sufficient for trajectory-level fidelity in the within-family Gamma validation studied here. However, later cross-family comparisons in Chapter 6 require finer resolutions, since moderate grids can still underestimate differences between Gamma and Lognormal trajectories.

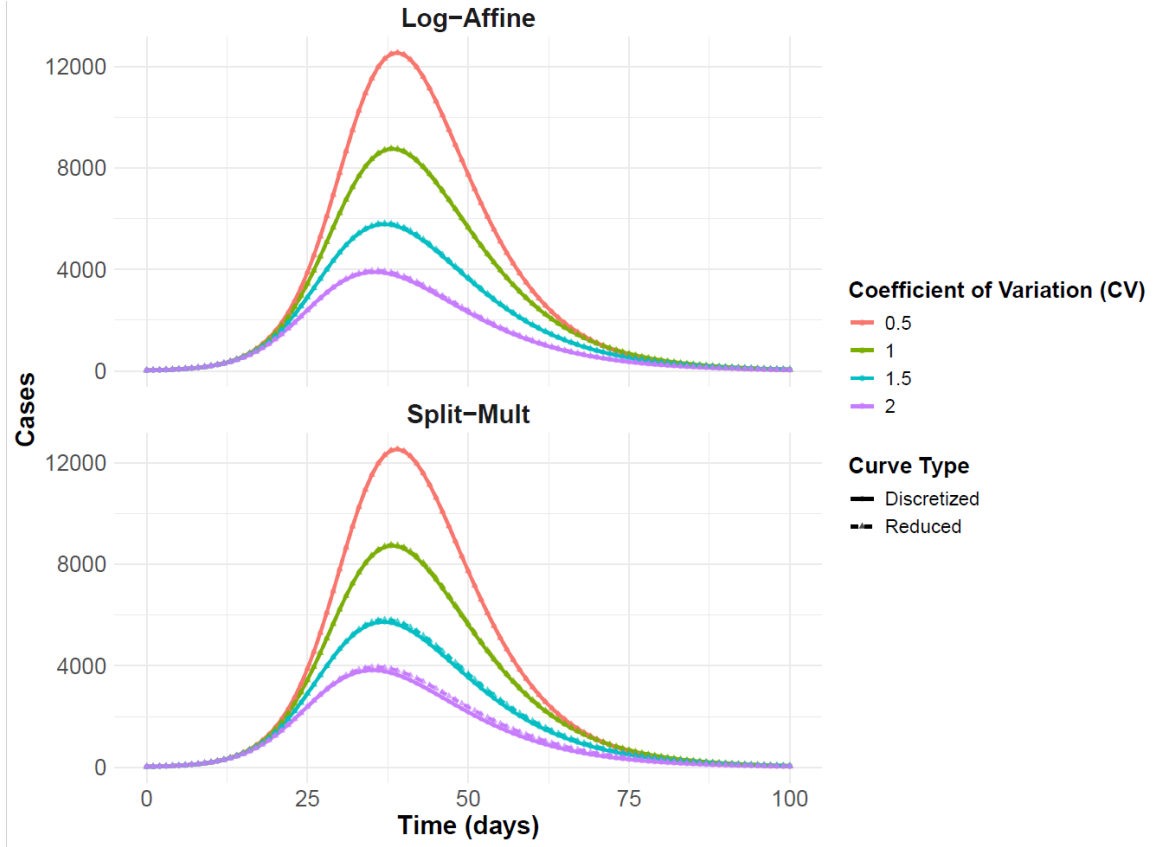


Figure 4.7: LA and SM Trajectories of discretised (*solid*) vs. reduced (*dashed*) infection curves at $k = 20$, split by transformation and coloured by CV. Agreement is visually exact across all CV values and both transformations.

4.6 Chapter Summary and Conclusion

4.6.1 Summary

The first main result of the chapter is methodological. A common three-stage pipeline was developed for positive susceptibility distributions: CDF binning to define the group probabilities, conditional-mean representatives to define the grid points, and a final transformation step to match the target mean and coefficient of variation. Two transformations were studied. The split-multiplicative (**SM**) approach applies separate rescalings to the lower and upper halves of the grid, whereas the log-affine (**LA**) approach uses the monotone map

$$x' = \exp(A + B \log x),$$

with the variance target enforced through the scalar equation $g(B) = 0$.

The second main result is that, in the Gamma case, the discretised model reproduces the reduced heterogeneous-susceptibility system very closely once the number of groups is moderately large. The trajectory-matching experiments in Section 4.5 showed that errors decrease rapidly as

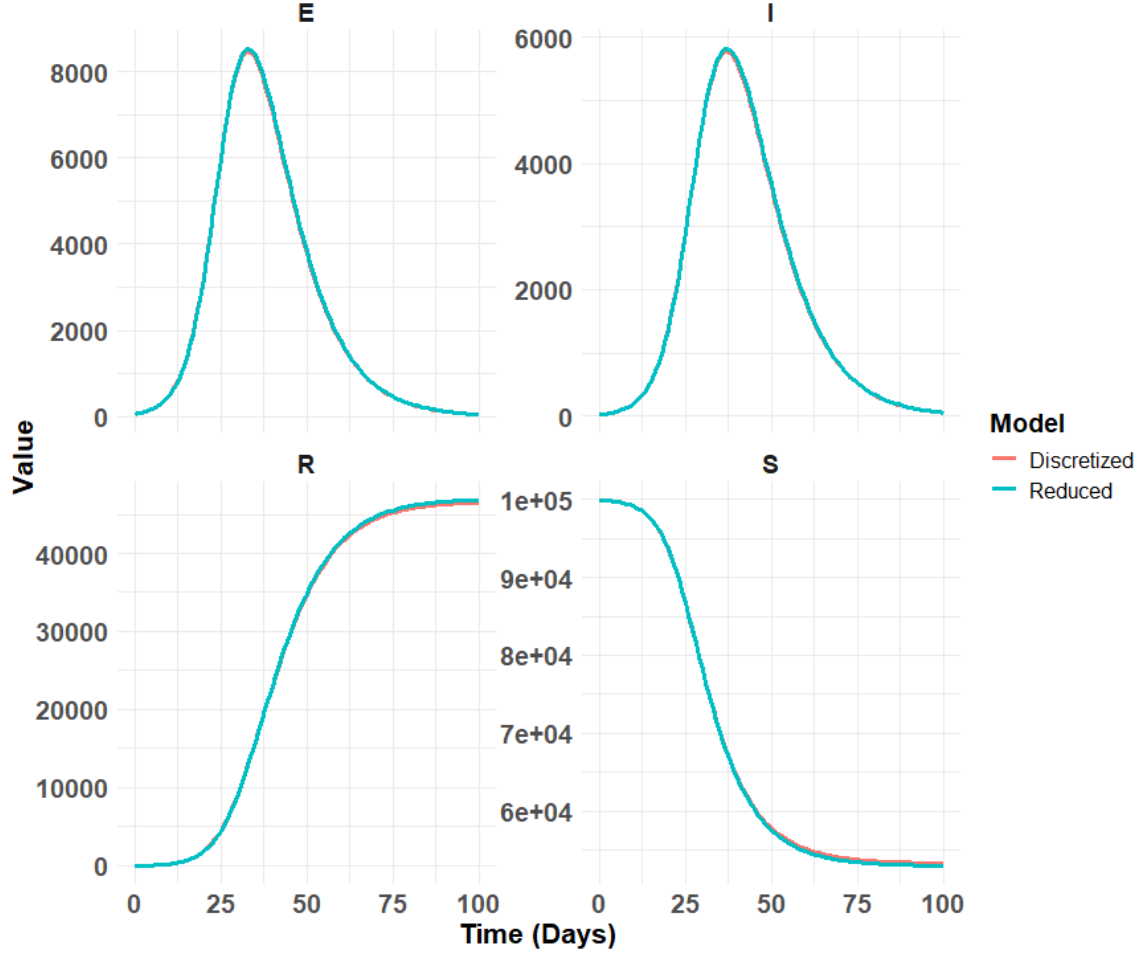


Figure 4.8: Trajectory comparison at moderate heterogeneity ($CV \approx 1.5$, $\alpha = 0.44$) with the log-affine transformation. The discrete-group model (red) tracks the reduced Gamma model (blue) almost exactly. Even with increased variance in susceptibility, no visible differences in the epidemic curves are observed.

k increases, with the largest improvement occurring between $k = 10$ and $k = 20$. Peak timing differences were generally 0 to 1 day, and for $k = 20$ the discretised and reduced trajectories were already in very close agreement across the heterogeneity levels considered. Where both transformations were admissible, LA and SM gave very similar trajectory-level results.

The third result is that, although both methods perform well in moderate settings, **LA** is the more **robust** (and therefore recommended method) of the two. In particular, the **SM** quadratic can fail to produce an admissible positive solution on coarse grids at high heterogeneity, whereas **LA** remains stable and continues to provide a valid calibrated grid. This makes **LA** the safer default transformation when the coefficient of variation is large or when the discretisation is relatively coarse.

4.6.2 Additional results from Appendix E

Appendix E extends the same discretisation pipeline from Gamma to the Lognormal family and provides the distribution-level checks that are not shown here. It shows that the same three-stage construction can be applied to mean-one Lognormal targets, with conditional-mean representatives and either **LA** or **SM** calibration preserving positivity and matching the target moments.

The appendix also adds a practical message about spacing choice (CDF binning). Distribution-level validation there shows that linear spacing performs better than equal spacing across the tested settings, and that this advantage is especially clear for the Lognormal family, whose heavier right tail is harder to capture accurately. In a nutshell, **LA** combined with linear spacing gives the lowest density errors across the tested configurations.

The appendix further shows that increasing the number of groups improves the distributional approximation, but the interpretation depends on the modelling aim. For within-family purposes such as moment matching and Gamma trajectory reproduction, moderate group sizes already work well. For inter-family comparisons, however, the grid must be finer. This point becomes important in Chapter 5, where the Gamma–Lognormal comparison is shown to be qualitatively visible at $k = 20$ but still underestimated in magnitude, while the comparison stabilises more reliably on finer grids such as $k \geq 40$.

4.6.3 Conclusion

In conclusion, this chapter established a workable discretisation framework for continuous susceptibility distributions in the heterogeneous-susceptibility SEIR model defined by equations (1.64)–(1.67). It showed that the framework is flexible enough to be distribution-agnostic, accurate enough to recover the Gamma benchmark closely, and stable enough to support later robustness analysis beyond the reduced closed-form setting.

The appendix results strengthen this conclusion by showing that the same framework extends naturally to Lognormal susceptibility and by identifying the combinations that work best in practice, especially **LA** with linear spacing. Therefore, the chapter provides the numerical bridge needed to move from the reduced Gamma model used in Chapters 2 and 3 to the cross-family robustness and misspecification analysis in Chapter 5.

Chapter 5

Distribution Misspecification in Heterogeneous Susceptibility Models

5.1 Introduction

In Chapter 4 we motivated and developed a method for discretising the continuous susceptibility distribution and simulating the resulting finite susceptibility-stratified heterogeneous SEIR model (4.3). The scheme preserves the unit mean and a prescribed coefficient of variation for a given distribution, and was tested on the Gamma and Lognormal families. In this Chapter we apply that discretisation framework to revisit the combined roles of susceptibility heterogeneity and non-pharmaceutical interventions (NPIs) studied in Chapters 2 and 3, but now using the finite susceptibility-stratified system (4.3) rather than the reduced Gamma system in equations (1.78)–(1.81), on which those chapters relied. Chapters 2 and 3 showed, for Gamma-distributed susceptibility, that a homogeneous model misattributes the effect of heterogeneity to NPIs, and that fitting two concurrent epidemics with shared parameters alleviates the resulting identifiability problems [1]. A natural question, left open by those chapters, is whether their conclusions depend on the assumed Gamma form.

While the presence of heterogeneity in susceptibility is widely acknowledged, its precise empirical distributional form in any given population remains uncertain, so a natural question is: *does the shape of the susceptibility distribution matter?* Rose et al. [43] argue that heterogeneity effectively “sets the order” of the system by shaping the force of infection (FOI): under selective depletion the Gamma distribution behaves as an “eigen-distribution”, preserving its functional form as highly susceptible individuals are infected and removed, and inducing a power-law relation between the FOI and the susceptible fraction whose exponent is fixed by the Gamma shape

parameter. This selective, accelerated removal of high-risk individuals progressively *sculpts* the susceptible pool, so that the composition of those who remain becomes increasingly dominated by individuals of lower susceptibility as the epidemic proceeds. Because the shape of the distribution controls how quickly this sculpting occurs (see, e.g., [4, 43]), two families that agree on their first two moments can still drive measurably different epidemic dynamics through their higher-order structure.

The consequences of structural misspecification in infectious-disease models have received growing attention across several modelling traditions, and they establish the concern that motivates this Chapter. In pairwise accelerated failure time (AFT) models of transmission, Sharker et al. [82] showed that an incorrect assumption about the contact-interval distribution can bias mechanistic inference even when the overall epidemic pattern is reproduced accurately: regression effects for susceptibility remained robust under moderate misspecification, but quantities tied directly to the transmission mechanism (hazard structure, transmission timing, secondary attack risk) were sensitive to the assumed family. In a complementary direction, Gao et al. [83] found that in Bayesian phylodynamic models the default priors on dispersal encode strong, biologically unrealistic assumptions that the sparse data cannot override, qualitatively distorting the inferred spatial history of outbreaks. A recurring pattern emerges from these studies: a misspecified model can reproduce the observed data by *compensating* through other parameters, creating a false sense of model adequacy. This risk is especially more serious in heterogeneous epidemic models, where selective depletion dynamically reshapes the effective susceptible population, so that two susceptibility distributions with identical first and second moments may still diverge through their tails, and where interventions such as NPIs offer an additional channel through which misspecification can be absorbed during inference.

The choice of the Gamma family as the default for susceptibility is not arbitrary. Beyond its eigen-distribution property, it permits the explicit infinite system to be reduced to a closed-form low-dimensional system of ordinary differential equations [44, 52, 84], making inference computationally tractable, and it has been used widely in applied work: to lower the herd-immunity threshold under individual variation [4, 44], to interpret vaccine efficacy through frailty-mixing models [84–87], and to analyse dose–response and challenge experiments in which host susceptibility is inferred from infectivity–dilution curves [88–93].

This breadth of use makes the Gamma family a sensible benchmark, but also makes it important to understand what is at stake if the data-generating distribution differs from it.

The Gamma distribution is, however, not the only candidate, and the empirical form in a real

population is unknown. Several alternatives have been proposed, and it is useful to ask which of them can even be placed on the common footing used throughout this thesis, namely a fixed unit mean ($\mu = 1$) with heterogeneity carried entirely by the variance $V = \nu^2$. Fixing the mean standardises the location scale and allows tail thickness and skewness to be compared without confounding. Under this constraint the families divide cleanly into those that admit a unit-mean parameterisation and those that do not:

- **Gamma**, **Lognormal** and the **Inverse Gaussian** distributions are all supported on $(0, \infty)$ and admit a unique unit-mean reparameterisation in terms of the variance. For the Gamma, shape $\alpha = 1/V$ and scale $\theta = V$; for the Lognormal, log-scale variance $\sigma^2 = \ln(1 + V)$ and log-location $\mu_L = -\frac{1}{2} \ln(1 + V)$; for the Inverse Gaussian, location $\mu = 1$ and shape $\lambda = 1/V$. All three are right-skewed, with the Lognormal [94] carrying a heavier right tail than the Gamma at matched moments, and the Inverse Gaussian arising naturally from frailty theory [95]. As a compact reminder, for unit mean the coefficient of variation relates to the native parameters as $\nu = 1/\sqrt{\alpha}$ for the Gamma and $\nu = \sqrt{e^{\sigma^2} - 1}$ for the Lognormal.
- The **Beta** distribution has been proposed as a physically realistic bounded alternative because it confines susceptibility to a finite domain and so avoids the unbounded “super-susceptibility” tail of the Gamma or Lognormal [52,86,87,89,92,96,97]. In its standard form on $[0, 1]$, however, its mean $\alpha/(\alpha + \beta)$ is strictly less than one for any proper density, and forcing the mean to one requires $\beta \rightarrow 0$, which collapses the distribution to a degenerate point mass at $x = 1$ (zero variance). A unit-mean, non-degenerate Beta model therefore requires rescaling the support to $[0, b]$ with $b > 1$, at which point the family can even be left-skewed rather than right-skewed.
- The **Positive Stable** distribution, used to model power-law transmission and heavy-tailed frailty [25,98,99], cannot be given a finite unit mean at all: for stability index $\alpha \leq 1$ its right tail decays as $x^{-(1+\alpha)}$ and the first moment diverges, while $\alpha > 1$ forces the support onto the whole real line, violating positivity.

Consequently, of the commonly cited families only the Gamma, Lognormal and Inverse Gaussian sit naturally within the unit-mean framework of this thesis.

Because our aim is to separate the effect of distributional *shape* at matched mean and variance, and because the Gamma and Lognormal are by far the two most frequently adopted positive families [4,94], the numerical experiments in this Chapter use *only* the Gamma and Lognormal

distributions. The Beta, Inverse Gaussian and Positive Stable families are introduced here to situate that modelling choice, and to make explicit which alternatives are and are not admissible under a unit mean; they are not simulated.

Although we do not simulate the other families, the selective-depletion mechanism identified above lets us reason about how they would be expected to behave, which is instructive for interpreting the Gamma–Lognormal comparison. The operative lever is the *right tail*: a reservoir of highly susceptible individuals sustains transmission chains longer, delays the point at which sculpting brakes the epidemic, and so produces a larger and later peak. On this reasoning, a family with a *heavier* right tail than the Lognormal at the same ν (for instance the Inverse Gaussian at high variance, or a near-power-law form) would be expected to push the peak still higher and later than the Lognormal, whereas a family *without* a heavy right tail would move in the opposite direction. Distributions that are bounded above, symmetric, or left-skewed—for example a rescaled Beta on $[0, b]$, a truncated normal centred at one, a uniform distribution on $(0, 2)$, or a Weibull with large shape parameter, all of which admit a unit mean—lack the super-susceptible reservoir and would therefore brake the epidemic less abruptly, behaving closer to the homogeneous model. Since the homogeneous case ($\nu = 0$) is the extreme of no heterogeneity, and Chapters 2 and 3 showed that it substantially *over*-predicts epidemic size, a non-right-skewed susceptibility family would be expected to yield forecasts lying between the Gamma prediction and the (more pessimistic-to-optimistic) homogeneous prediction, rather than beyond the Lognormal. These are mechanistic expectations rather than simulated results, but they frame the Gamma and Lognormal as spanning a representative, and arguably the most consequential, portion of the plausible range of right-skewed susceptibility shapes.

The Gamma and Lognormal families that we do compare are constructed to share the same first and second moments, so any remaining differences are attributable to their higher-order structure. Figure 5.1 illustrates the two densities for a unit mean and $CV \in \{0.5, 1, 1.5, 2\}$: for $CV < 1$ the Gamma and Lognormal have similar shapes, whereas for $CV > 1$ the Lognormal peaks higher and is more strongly right-skewed. While both families are matched on their first two moments by construction, their higher-order characteristics differ substantially: Figure 5.2 shows that the Lognormal has higher skewness and kurtosis, increasingly so at larger coefficients of variation. These theoretical differences translate into measurable differences in epidemic trajectories and forecasts, as we demonstrate through the rest of the Chapter.

We investigate the consequences of this uncertainty through two primary questions:

1. **Trajectory comparison with fixed mean and coefficient of variation.** What

are the differences between Gamma- and Lognormal-based heterogeneous SEIR epidemic trajectories when the two are matched on mean and CV and interventions are absent? We compare infection incidence, together with peak timing, peak magnitude, and final size, across a grid of CV values.

2. **Inference and forecast under misspecification.** What are the consequences of misspecifying the susceptibility distribution, under the combined effects of heterogeneity and NPIs, for single and joint inference from epidemic time-series data, and for the resulting forecasts? We quantify estimation uncertainty in the key parameters $(\mathcal{R}_0, \nu, t_0, c^*)$ (as in Chapters 2 and 3), and assess how joint two-epidemic fitting sharpens *identifiability and forecasts* while potentially amplifying structural bias. Throughout, each dataset is fitted with three candidate models—a correctly specified heterogeneous model, a misspecified heterogeneous model, and a homogeneous reference model ($\nu = 0$)—so that the cost of choosing the wrong family can be judged against the cost of ignoring heterogeneity altogether.

The Chapter is organised as follows. Section 5.2 recalls the susceptibility-stratified SEIR framework and the moment-matched discretisation used to place Gamma and Lognormal models on common ground (matching mean = 1 and coefficient of variation ν). Section 5.3 addresses Question 1, quantifying the deterministic trajectory divergence that arises purely from distributional shape (peak magnitude, peak timing, and final size) across a grid of heterogeneity levels and discretisation resolutions. Section 5.4 describes the misspecification experiment for Question 2, examining single-epidemic inference, the identifiability gains from two-epidemic fitting, and how those gains can translate into *precise but biased* inference under structural misspecification, including fit-and-forecast demonstrations under joint inference for the **Lognormal baseline**. Finally, Appendix F reports the complementary **Gamma-baseline** experiments (single and two epidemics) and closes with a forecast-based diagnostic.

5.2 Trajectory comparison and analysis

We compare simulated epidemic trajectories without any observational noise and without effect of NPIs. This Chapter uses the finite susceptibility-stratified heterogeneous SEIR model (4.3) developed in Chapter 4.

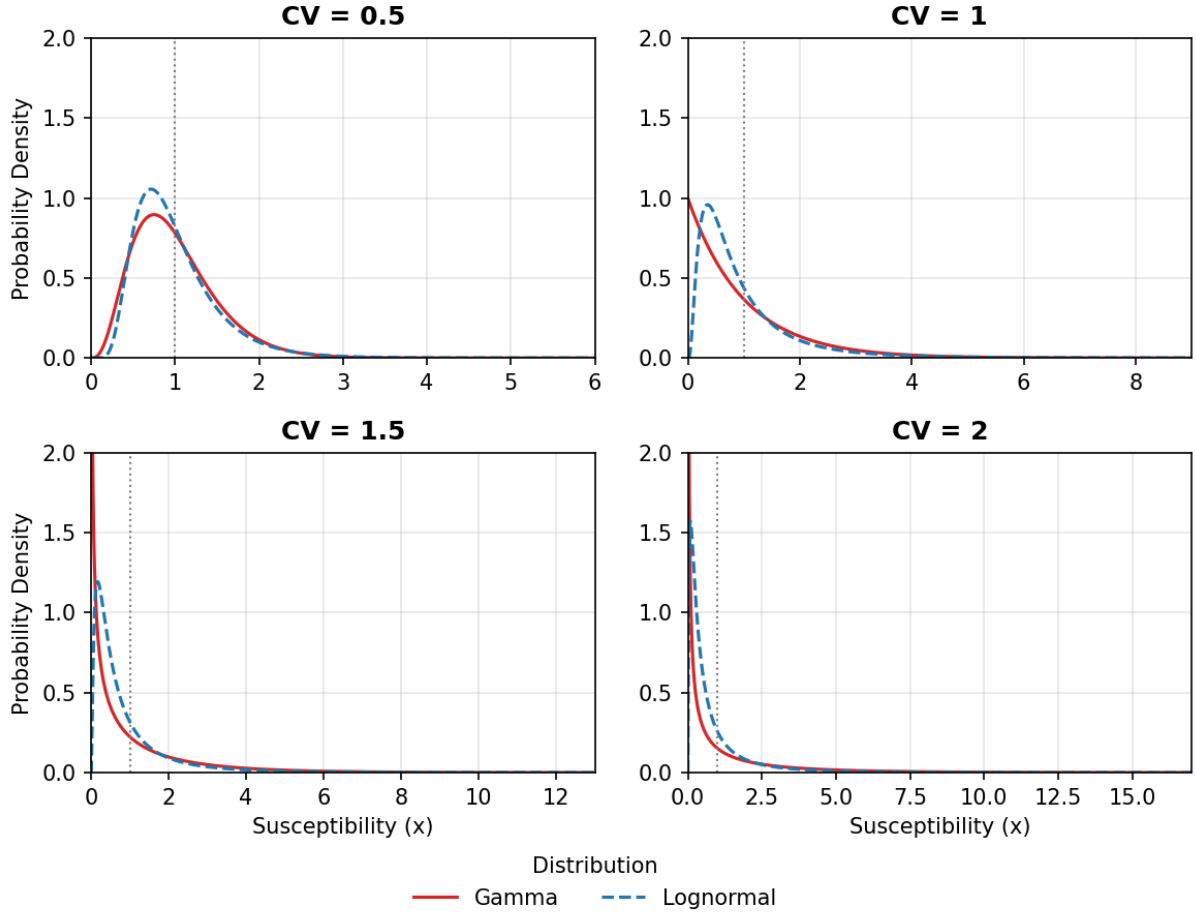


Figure 5.1: Theoretical distribution comparison between Gamma and Lognormal distributions with identical means ($= 1$) and coefficients of variation. With matching first and second moments, the distributions reveal different probability density shapes, particularly in tail behaviour. The differences become more pronounced at higher coefficients of variation, with the Lognormal distribution showing heavier tails and more probability mass at extreme susceptibility values.

5.2.1 Transmission model used in this chapter

This chapter uses the finite susceptibility-stratified heterogeneous SEIR model in equation (4.3), with coefficient of variation defined in equation (5.1). We therefore do not repeat the full governing equations here. The key object for the present analysis is the finite susceptibility grid

$$\{(q_i, x_i)\}_{i=1}^k,$$

where q_i is the population fraction assigned to susceptibility group i , and x_i is the representative susceptibility of that group. The corresponding state variables are $S_i(t)$, $E_i(t)$, $I_i(t)$, and $R_i(t)$.

Each grid is constructed to match unit mean and a prescribed coefficient of variation ν :

$$\nu = \sqrt{\sum_{i=1}^k q_i (x_i - 1)^2}, \quad \sum_{i=1}^k q_i x_i = 1. \quad (5.1)$$

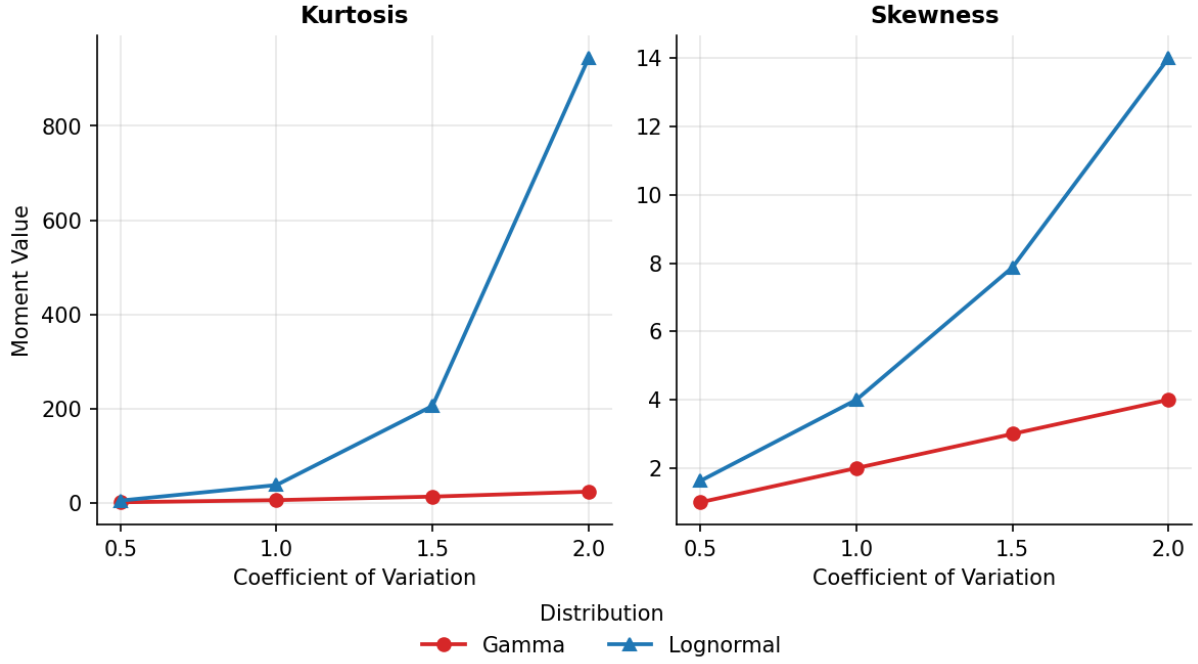


Figure 5.2: Higher-order moment differences between Gamma and Lognormal distributions. With both distributions having identical means and variances by construction, their third and fourth moments (skewness and kurtosis) differ markedly. The Lognormal distribution shows exponentially increasing skewness and kurtosis with the coefficient of variation, indicating fundamentally different tail behaviours that may influence epidemic dynamics during the pre-convergence period.

Thus Gamma and Lognormal susceptibility distributions are compared at the same first two moments. For a mean-one Gamma distribution $X \sim \text{Gamma}(\alpha, \alpha)$ under the shape-rate convention, $\nu = 1/\sqrt{\alpha}$, while for a unit-mean Lognormal distribution, $\nu = \sqrt{\exp(\sigma^2) - 1}$. Any remaining differences between the two models therefore arise from distributional shape, especially higher-order moments and tail behaviour.

The basic reproduction number is defined as in Equation (1.69). Because susceptibility is normalised to have mean one, heterogeneity does not change the initial expression for $\mathcal{R}_0$; its effect enters during the epidemic through selective depletion of the susceptible pool.

For the deterministic trajectory comparisons in this section, cumulative onset incidence is tracked within each susceptibility group through

$$\frac{dC_i}{dt} = \delta E_i(t),$$

and the total simulated incidence is obtained by summing over groups. Therefore, differences between Gamma and Lognormal trajectories in this section are attributable to the susceptibility

distribution rather than to time-varying transmission control.

5.2.2 Discretisation and model initialization

To compare trajectories fairly, the discrete susceptibility distributions $\{(q_i, x_i)\}_{i=1}^k$ for both Gamma and Lognormal models are constructed using the log-affine discretisation method developed in Section 4.2. We use equal-probability bins and conditional-expectation representatives, followed by the log-affine moment correction so that both distributions match the target mean $\mu = 1$ and coefficient of variation ν . Therefore, any remaining differences between the simulated trajectories are attributable to distributional shape, especially higher-order moments and tail behaviour.

5.2.3 Trajectory comparison metrics

We compare two *discretised* heterogeneous susceptibility SEIR systems (4.3) that differ only in the family used for the susceptibility distribution: (i) **Lognormal** and (ii) **Gamma**. Both are initialised with identical SEIR model parameters:

1. Total population size N and basic reproduction number (1.69)
2. Let E_0 and I_0 be the total numbers initially exposed and infectious. We set initial conditions (E_0, I_0) such that $E_0 \approx 2.5 I_0$ with justification in section A.1.3 of Appendix A.
3. The initial conditions (E_0, I_0) are then distributed among k susceptibility classes proportionally to q_i . For $i = 1, \dots, k$, we initialise:

$$E_i(0) \approx E_0 q_i, \quad I_i(0) \approx I_0 q_i, \quad S_i(0) = (N - E_0 - I_0) q_i, \quad R_i(0) = 0,$$

The values of fixed parameters of the model equations (4.3) as well as the values of the (E_0, I_0) are those from Chapter 3 in Table 2.1. We use incidence as the primary trajectory quantity for addressing Question 1. Let $C_{\mathcal{D}}(t)$ denote the total cumulative incidence, summed over all susceptibility groups, under susceptibility distribution

$$\mathcal{D} \in \{\Gamma, \text{LN}\},$$

where Γ denotes the Gamma distribution and LN denotes the Lognormal distribution. The

corresponding daily incidence is

$$y_{\mathcal{D}}(t) = \Delta C_{\mathcal{D}}(t) = C_{\mathcal{D}}(t) - C_{\mathcal{D}}(t-1), \quad y_{\mathcal{D}}(0) = 0. \quad (5.2)$$

We compare the Gamma and Lognormal models using the following summary measures:

1. Relative difference in peak incidence magnitude (percent):

$$\Delta y_{\text{peak}}^{(\%)} = 100 \frac{\max_t y_{\text{LN}}(t) - \max_t y_{\Gamma}(t)}{\max_t y_{\Gamma}(t)}. \quad (5.3)$$

2. Difference in peak-incidence timing (days):

$$\Delta t_{\text{peak}} = t_{\text{LN}}^{\max} - t_{\Gamma}^{\max}, \quad t_{\mathcal{D}}^{\max} = \arg \max_t y_{\mathcal{D}}(t). \quad (5.4)$$

3. Relative difference in final epidemic size (percent):

$$\Delta C_T^{(\%)} = 100 \frac{C_{\text{LN}}(T) - C_{\Gamma}(T)}{C_{\Gamma}(T)}, \quad (5.5)$$

where T denotes the simulation horizon.

Remark

We take Gamma as reference distribution: a baseline in this thesis because the reduced heterogeneous model used in earlier chapters is derived under Gamma-distributed susceptibility. We report differences as $(\text{LN} - \Gamma)$, taking the Gamma model as the reference trajectory. This is supported by the fact that the Gamma family is a natural benchmark for susceptibility heterogeneity: under contagion-driven selective depletion it is an eigen-distribution, and other initial distributions are often sculpted toward Gamma-like forms [43].

5.2.4 Simulation grid

We vary heterogeneity $\nu \in \{0.5, 1, 1.414, 2\}$ and discretisation resolution $k \in \{10, 20, 40, 60, 100\}$. For each (ν, k) pair we simulate the Gamma and Lognormal models matched on mean and CV via section 5.2.2, and summarise differences using the metrics in section 5.2.3.

5.3 Trajectory comparison of Gamma vs Lognormal distributions using Fixed mean and CV

Our first result is to establish whether the *distributional shape* of susceptibility (Gamma vs. Lognormal), holding the first two moments equal, induces systematic differences in epidemic trajectories. We examine these differences across varying levels of CV and discretisation resolution (k).

5.3.1 The pattern of divergence of trajectories

When heterogeneity is low ($\nu \leq 0.5$), the epidemic curves (as simulated from system (4.3)) for Gamma and Lognormal distributions are nearly indistinguishable. At low ν , both distributions appear roughly symmetric around the mean, producing overlapping epidemic curves with near-zero difference. However, as heterogeneity increases, a distinct behavioural divergence emerges.

Figure 5.3 illustrates the divergence between the trajectories of the two distributions using a high-resolution discretisation ($k = 100$). As ν increases from 0.5 to 2.0, the shape of the susceptibility distribution becomes a dominant driver of epidemic dynamics. A clear pattern is visible:

1. **Incidence magnitude:** The Lognormal model (blue) produces a higher peak incidence than the Gamma model (red) at every value of ν .
2. **Timing:** The Lognormal epidemic peaks later than the Gamma epidemic (Table 5.1).
3. **Relative excess:** The Lognormal-to-Gamma peak ratio rises monotonically with ν , from 1.010 at $\nu = 0.5$ to 1.311 at $\nu = 2.0$ ($k = 100$; Table 5.1).

Mechanistically, the Gamma distribution provides a stronger heterogeneity braking effect for the same ν . It depletes its high-susceptibility core rapidly, reducing the effective reproduction number earlier. In contrast, the Lognormal distribution possesses a heavier right tail at the same ν , containing more highly susceptible individuals who sustain transmission chains longer, preventing the epidemic from slowing down as quickly. Consequently, we characterise the **Gamma model as optimistic** (predicting a smaller, earlier epidemic) and the **Lognormal model as pessimistic** (predicting a larger, protracted epidemic) when both are conditioned on the same mean and ν . At $\nu = 2.0$, this choice of distribution alone changes the cumulative incidence by nearly 50%, as indicated in Figure 5.4.

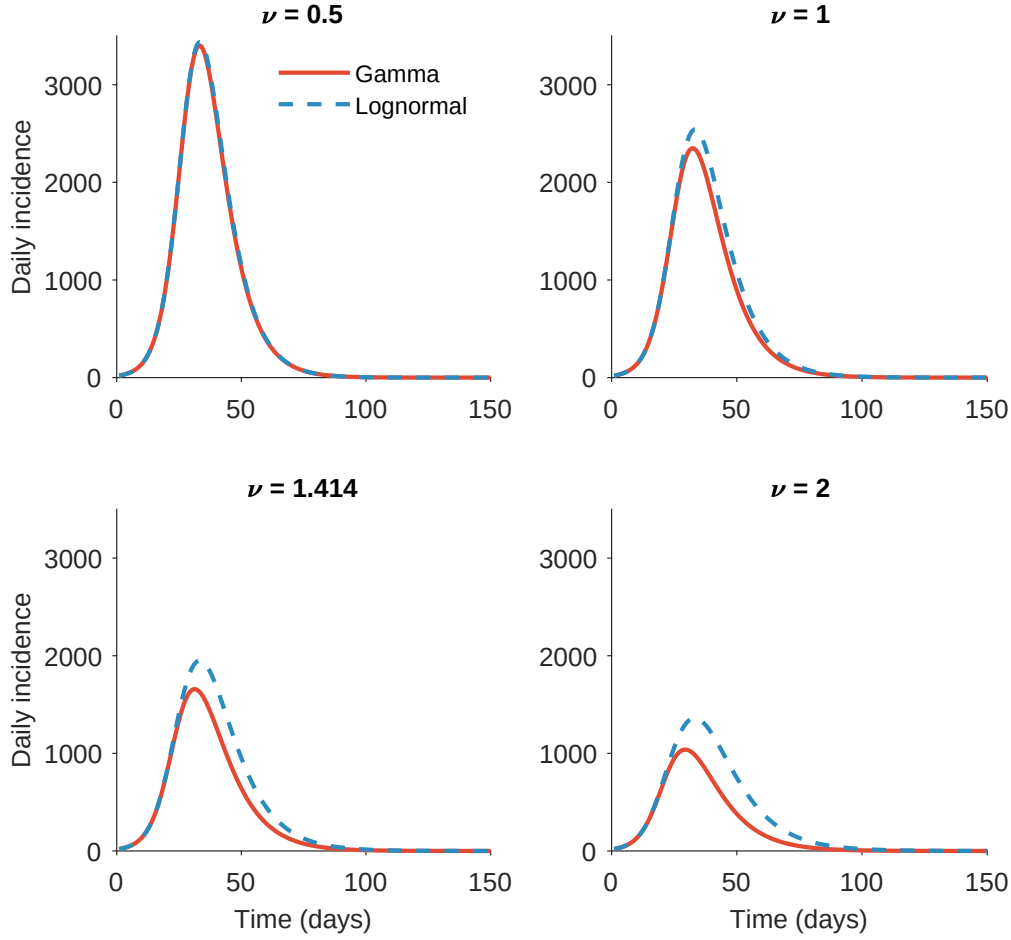


Figure 5.3: Trajectory comparison at high discretisation resolution ($k = 100$). Daily incidence from Gamma-discretised (red) and Lognormal-discretised (blue) heterogeneous SEIR models for four levels of heterogeneity ($\nu \in \{0.5, 1, 1.414, 2\}$). Both distributions are matched on mean susceptibility ($= 1$) and coefficient of variation; no NPIs are applied. At low heterogeneity ($\nu = 0.5$) the two distributions produce nearly identical epidemic curves. As ν increases, the Lognormal model produces a progressively higher and later peak. All four panels share the same vertical scale; see Table 5.1 for numerical summaries.

5.3.2 The impact of discretisation (k) and convergence

The simulation experiments further reveal that although Chapter 4 recommends a discretisation resolution as low as $k = 20$ as sufficient for modelling system (4.3) within a single family (Gamma alone), this resolution can mask structural differences when Gamma and Lognormal distributions are compared.

Figure 5.5 shows the same comparison with $k = 20$. While the general pattern holds, the magnitude of the difference is underestimated relative to $k = 100$. Table 5.1 makes this explicit: at $\nu = 1.414$ the peak incidence difference is $+292.9$ cases day^{-1} ($+17.66\%$) at $k = 100$, but only $+164.8$ cases day^{-1} ($+10.28\%$) at $k = 20$, a factor of 1.78. At $\nu = 2.0$ the absolute

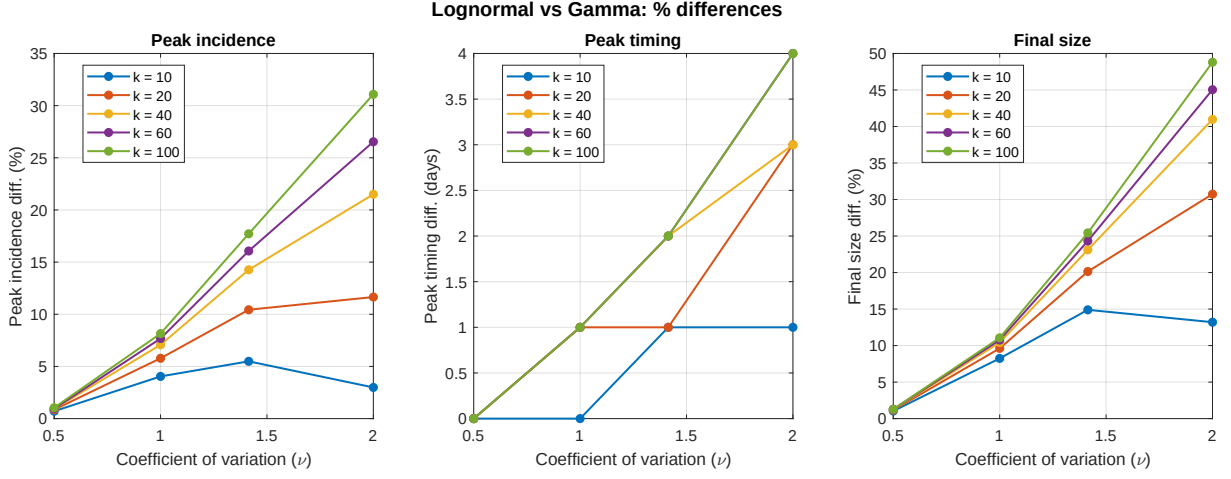


Figure 5.4: Epidemic outcome differences between Lognormal and Gamma susceptibility distributions increase with heterogeneity and depend on discretisation resolution. Each panel shows the percent difference $100 \times (\text{Lognormal} - \text{Gamma}) / \text{Gamma}$ across $\nu \in \{0.5, 1.0, 1.414, 2.0\}$ for five resolutions $k \in \{10, 20, 40, 60, 100\}$: peak incidence (left), peak timing (middle), and final size (right). All differences are non-negative: the Lognormal consistently produces equal or greater values than the matched Gamma. At coarse resolution ($k = 10$), differences plateau at high ν , underestimating the distributional divergence observed at high resolution. At $k \geq 40$, the results stabilise, revealing an approximately linear increase with ν across all three outcomes.

difference actually falls from +322.8 ($k = 100$) to +112.5 ($k = 20$), even though the ratio remains similar (1.311 versus 1.116), because the coarse grid compresses the heavy Lognormal tail and underestimates the data-generating peak.

Figure 5.4 makes the convergence pattern explicit across all three outcome metrics and the full ν grid. At low heterogeneity ($\nu = 0.5$) results are stable across all k : peak incidence differences range from 0.71% to 1.05% and final size differences from 1.04% to 1.30%, indicating that coarse discretisation is adequate when distributional shape differences are inherently small. However, the picture changes at high heterogeneity. At $\nu = 2.0$, the coarsest grid ($k = 10$) returns a peak incidence difference of only 2.99%, a final size difference of 13.20%, a factor of approximately 10 and 3.7 lower, respectively, than the estimates obtained at $k = 100$ (31.09% and 48.80%). When ν is large, the Lognormal distribution has a heavy right tail, and a coarse discretisation effectively compresses that tail into too few extreme groups, making Gamma and Lognormal look artificially similar and potentially masking a real structural difference.

Convergence across k is effectively achieved by $k \geq 40$, where the curves in Figure 5.4 cluster tightly and the three metrics change by at most 3–4 percentage points relative to $k = 100$ across all ν values. We therefore treat $k \geq 40$ as producing a reliable estimate of the cross-family difference and use $k = 100$ as a conservative upper reference throughout.

Thus, even when Gamma and Lognormal models are matched exactly on mean and coefficient of variation, higher-order distributional structure produces a measurable deterministic signal

in incidence trajectories once heterogeneity is moderate to high. In particular, Lognormal susceptibility yields a later and larger incidence peak and a larger final size than Gamma. These deterministic differences provide the baseline signal for the next section, where we quantify how such shape effects translate into parameter bias under misspecification.

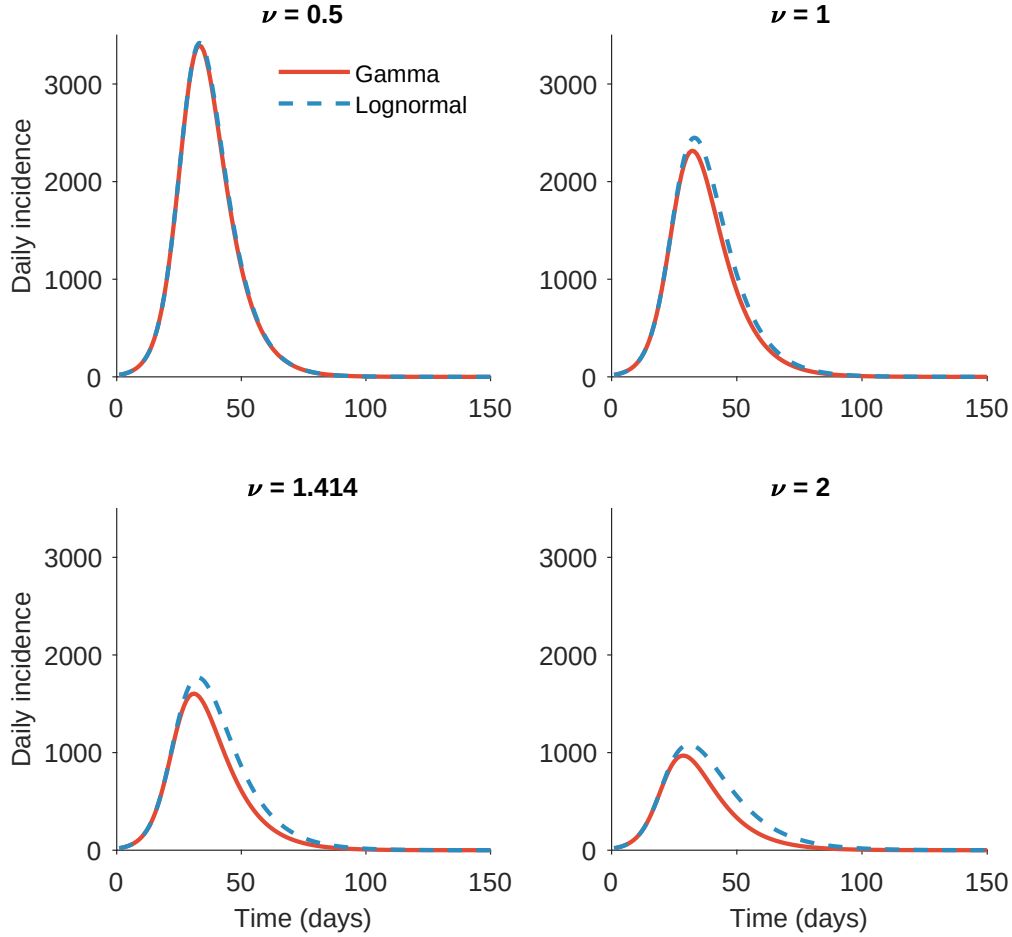


Figure 5.5: Trajectory comparison at moderate discretisation resolution ($k = 20$). Same parameters as Figure 5.3 but using $k = 20$ susceptibility groups. The qualitative pattern is preserved: the Lognormal model produces a larger, later epidemic than the Gamma model at each ν . However, the magnitude of the difference is underestimated relative to $k = 100$ (Table 5.1). All four panels share the same vertical scale as Figure 5.3, enabling direct visual comparison across resolutions.

Table 5.1: Peak incidence and timing: Gamma versus Lognormal susceptibility at $k = 100$ and $k = 20$. No-intervention experiment ($\mathcal{R}_0 = 3$, $N = 100,000$, $E_0 = 100$, $I_0 = 40$, $t_{\max} = 150$ days). Absolute difference and percentage difference are Lognormal minus Gamma; ratio is Lognormal peak divided by Gamma peak. The ratio column rises monotonically with ν at both resolutions, showing that the *relative* Lognormal excess grows with heterogeneity even though the absolute peaks fall. The absolute difference at $k = 20$ is smaller than at $k = 100$ at every $\nu \geq 1$, and reverses in ordering at $\nu = 2.0$, confirming that $k = 20$ underestimates the cross-family divergence at high heterogeneity.

ν	k	Peak Gamma	Peak LN	Peak day (G / LN)	Abs. diff.	% diff.	Ratio
0.500	100	3403.9	3437.1	33 / 33	+33.2	+0.98	1.010
	20	3394.9	3422.2	33 / 33	+27.3	+0.80	1.008
1.000	100	2349.5	2541.9	32 / 33	+192.4	+8.19	1.082
	20	2316.0	2451.1	32 / 33	+135.1	+5.84	1.058
1.414	100	1658.1	1950.9	31 / 33	+292.9	+17.66	1.177
	20	1604.1	1769.0	31 / 33	+164.8	+10.28	1.103
2.000	100	1038.1	1360.9	29 / 33	+322.8	+31.10	1.311
	20	968.8	1081.3	29 / 31	+112.5	+11.61	1.116

5.4 Distribution Misspecification Analysis

Section 5.3 demonstrates that, at high heterogeneity, Gamma and Lognormal susceptibility specifications can generate noticeably different deterministic incidence trajectories when matched on the same mean and coefficient of variation. We now quantify how such structural differences manifest into statistical inference when the susceptibility family is misspecified. This section details the data generation process and the experimental design used to separate structural bias from observation noise.

5.4.1 Data generation and observation model

For a given correct specification (Gamma or Lognormal), we use system (4.3) to obtain the mean daily incidence. Let $E_i(t)$ denote the number of exposed individuals in susceptibility group i at day t . The model-implied mean daily incidence is the total flow from the exposed to the infectious compartment across all k groups,

$$x_t(\boldsymbol{\theta}) = \delta \sum_{i=1}^k E_i(t), \quad (5.6)$$

where δ is the rate of progression from exposed to infectious (mean latent period $1/\delta$). The observed incidence is generated via a Poisson sampling model,

$$y_t \sim \text{Poisson}(x_t(\boldsymbol{\theta})), \quad t = 1, \dots, T. \quad (5.7)$$

We repeat this procedure independently across 200 replicated datasets for each baseline scenario across the grid described in Section 5.4.3.

5.4.2 Inference target and likelihood

Inference is performed by maximum likelihood estimation. For both the Gamma- and Lognormal-discretised susceptibility models described in Section 5.2.2, we use resolution $k = 100$ and estimate the parameter vector

$$\boldsymbol{\theta} := (\mathcal{R}_0, \nu, t_0, c^*). \quad (5.8)$$

As in Chapter 2 (Section 2.5), we assume the observations y_t are conditionally independent given the model states x_t . The likelihood function is

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{t=1}^T f(y_t \mid x_t(\boldsymbol{\theta})),$$

where $f(\cdot)$ is the Poisson probability mass function and $x_t(\boldsymbol{\theta})$ is the model-implied daily incidence from equation (5.6). The log-likelihood is therefore

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^T \log f(y_t \mid x_t(\boldsymbol{\theta})) \quad (5.9)$$

where the second equality follows from the Poisson mass function (dropping terms not depending on $\boldsymbol{\theta}$). For single-epidemic inference this likelihood is evaluated over one trajectory ($T = 100$); for joint two-epidemic inference the log-likelihoods of both trajectories are summed under the constraint that $\boldsymbol{\theta}$ is shared, as in Chapter 3, Section 3.3.1.

Because the analysis is based on replicated datasets, each numerical estimate reported in Section 5.5 and Appendix F is the median maximum-likelihood estimate across the valid replicates. Where interval limits are quoted, they are the means of the lower and upper endpoints of the replicate-specific 95% Wald confidence intervals. They summarise typical within-replicate uncertainty and should not be interpreted as a confidence interval for the across-replicate median.

5.4.3 Data-generating and fitted models

We adopt the same initial-condition specification (E_0, I_0) used in Chapters 2–3 for comparability across single- and two-epidemic designs.

Synthetic data are generated under one susceptibility family, which we refer to as the *baseline* distribution. The main chapter uses the **Lognormal baseline**: 200 replicate datasets are generated from the Lognormal-susceptibility model at $\nu = 1.414$ and the settings of Chapters 2

and 3. The complementary **Gamma baseline**, in which the directional roles of the two heterogeneous families are reversed, is reported in Appendix F.

Each baseline dataset is then fitted with three candidate models:

1. a **correctly specified heterogeneous model** (same family as the baseline);
2. a **misspecified heterogeneous model** (the alternative family); and
3. a **homogeneous reference model** obtained by fixing $\nu = 0$, which omits susceptibility heterogeneity entirely.

The homogeneous fit is included throughout so that the cost of choosing the wrong heterogeneous family can be judged against the cost of ignoring heterogeneity altogether, as studied under the Gamma assumption in Chapters 2 and 3. All fits use resolution $k = 100$.

5.5 Inference for Lognormal-distributed susceptibility

We begin the analysis with the scenario where the baseline population susceptibility follows a Lognormal distribution.

5.5.1 Single-epidemic inference: Lognormal baseline

Here we conduct inference in the standard scenario in which only one epidemic time series is available. Table 5.2 summarises the three fits across 200 replicated datasets at $\nu = 1.414$.

The two heterogeneous fits give similar estimates of $\mathcal{R}_0$, t_0 and c^* , all within 2.5% of the data-generating values. They differ on ν : the correctly specified Lognormal fit gives a median of 1.367 (relative bias -3.3%), while the misspecified Gamma fit gives 1.231 (relative bias -13.0%). Both heterogeneous fits have wide confidence intervals for ν (width 0.75 to 1.84) and sub-nominal coverage between 77% and 79%. The misspecification cannot be identified from a single epidemic by goodness-of-fit alone.

The homogeneous fit behaves qualitatively differently. It collapses c^* to a tight distribution at 0.245 (data-generating value 0.30, relative bias -18%) with 0% coverage, and biases $\mathcal{R}_0$ to 2.944 (-1.9%) with 19.5% coverage. The heterogeneity signal is absorbed into the intervention parameter, recovering the misattribution mechanism reported in Chapters 2 and 3 and in [1]. Figure 5.6 shows the homogeneous c^* density as a sharp peak well separated from both heterogeneous distributions.

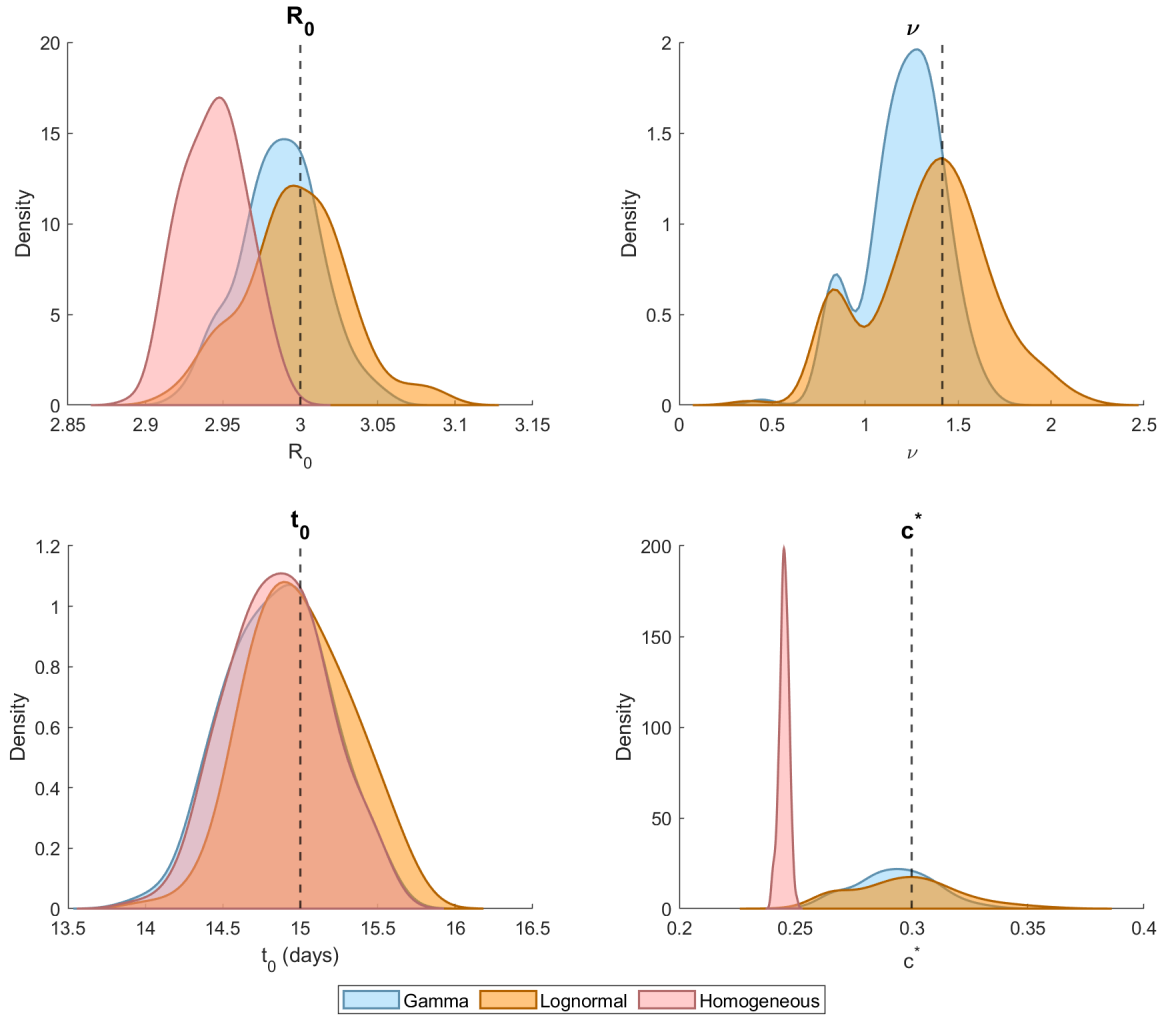


Figure 5.6: Distributions of single-epidemic MLE estimates for Lognormal-generated data ($\nu = 1.414$). Blue: Gamma (misspecified). Orange: Lognormal (correct). Pink: homogeneous (reference). Dashed vertical lines indicate the data-generating values. The two heterogeneous densities overlap substantially in R_0 , ν , t_0 and c^* . The homogeneous fit collapses c^* to a tight peak at 0.245.

Table 5.2: Single-epidemic inference for Lognormal-generated data ($\nu = 1.414$). Median (SD), mean 95% confidence-interval width, and empirical coverage across 200 replicates. Coverage is the percentage of replicates whose 95% confidence interval contains the data-generating value. Convergence (positive-definite Hessian): Gamma 183/200, Lognormal 176/200, homogeneous 200/200. The homogeneous fit fixes $\nu = 0$.

Model	Parameter	Median (SD)	CI width	Coverage (%)
Lognormal (correct)	$\mathcal{R}_0$	2.996 (0.034)	0.116	84.7
	ν	1.367 (0.323)	1.842	77.8
	t_0	14.986 (0.332)	1.341	94.9
	c^*	0.297 (0.022)	0.072	73.3
Gamma (misspecified)	$\mathcal{R}_0$	2.988 (0.026)	0.092	86.9
	ν	1.231 (0.208)	0.751	78.7
	t_0	14.867 (0.329)	1.269	96.2
	c^*	0.293 (0.017)	0.057	80.3
Homogeneous (reference)	$\mathcal{R}_0$	2.944 (0.020)	0.073	19.5
	ν	—	—	—
	t_0	14.871 (0.320)	1.169	92.5
	c^*	0.245 (0.002)	0.008	0.0

Note on ν coverage. Both heterogeneous fits show sub-nominal coverage for ν (correct Lognormal 77.8%, misspecified Gamma 78.7%). The two values are effectively indistinguishable: coverage reflects the joint effect of bias and interval width, and a single-epidemic likelihood surface that is nearly flat in the ν - c^* direction produces highly variable estimates whose intervals miss the data-generating value from both sides. Coverage alone therefore does not separate the correctly specified from the misspecified fit, which is a consequence of single-epidemic non-identifiability rather than evidence that misspecification is benign.

5.5.2 Fit and forecast for the Lognormal baseline (single epidemic)

From Section 5.5 it is clear that fitting a Gamma-susceptibility model to data generated by a Lognormal-susceptibility model infers a different coefficient of variation. This is not surprising, given the different shapes. The important question is whether the parametrised models retain useful predictive capacity despite the misspecification. To translate parameter estimates into forecasts, we lift NPIs at day $T = 100$ and simulate projected incidences to day 250 using each replicate’s fitted parameters.

Figure 5.7 shows the resulting trajectories. The two heterogeneous fits produce mean forecast trajectories that lie close to the deterministic baseline across the entire forecast window. Median peak height differs from the baseline by +2.9% for the correct Lognormal fit and by −6.5% for the misspecified Gamma fit; median final size differs by +1.7% and −12.5%, respectively. The homogeneous fit, by contrast, overshoots the peak by +216% and the final size by +88%, with no overlap between its 95% confidence band and the data-generating trajectory across most of the forecast window. The misspecified Gamma forecast error is therefore about 33 times smaller in magnitude than the homogeneous error in peak height and about 7 times smaller in final size.

In other words, the single-epidemic data, even when combined with the forecast period,

cannot separate the correctly specified from the misspecified heterogeneous fit. The mean Gamma trajectory points in the wrong direction, producing a smaller and earlier peak, but the wide uncertainty inherited from single-epidemic non-identifiability keeps its confidence band overlapping the data-generating trajectory. The cost of omitting heterogeneity altogether is an order of magnitude larger and is unambiguous even from a single epidemic. In the next section, we turn to two-epidemic joint inference, which is required to rule out the structurally incorrect heterogeneous model.

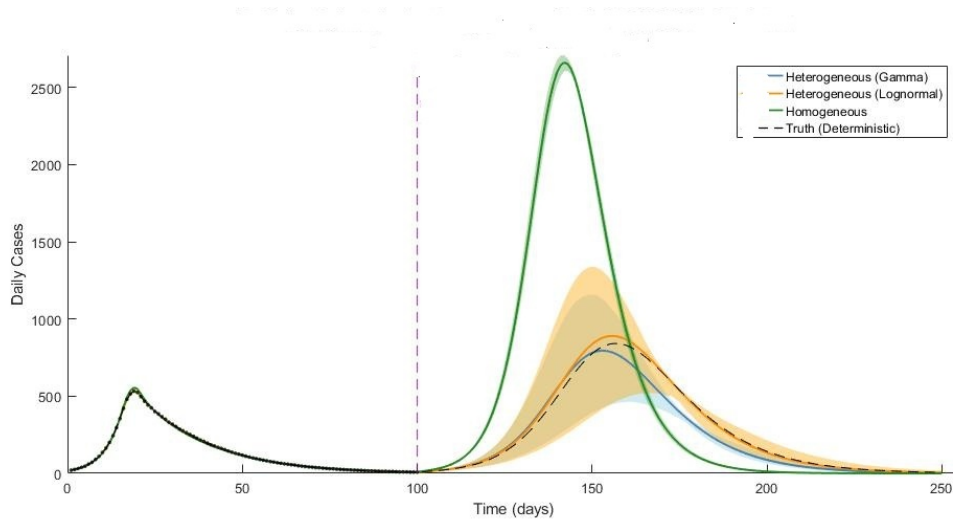


Figure 5.7: Single-epidemic forecast trajectories for Lognormal-generated data ($\nu = 1.414$, NPIs lifted at day 100). The black dashed line is the deterministic data-generating trajectory; shaded bands are 95% confidence envelopes across replicated fits. The Lognormal fit (orange) and the Gamma fit (blue) both lie close to the data-generating trajectory in their median trajectories, and their confidence bands overlap it. The homogeneous fit (green) produces a much higher and earlier peak.

5.5.3 Two-epidemic inference: Lognormal baseline

In Section 5.5.1 we found that misspecifying a Lognormal distribution as Gamma could not be diagnosed from a single epidemic: the confidence intervals were wide enough to absorb the bias, and the forecast confidence bands wide enough to absorb the divergence beyond the training window. This is consistent with the ν - c^* correlation documented in Chapter 2 and in [1], which supplies a compensation pathway that lets the optimiser trade heterogeneity against intervention strength. Following Chapter 3, we now conduct inference on pairs of epidemics with shared parameters to alleviate this correlation and narrow the confidence intervals.

We simulate pairs of epidemics with the same parameter values but different initial conditions (large seed: $E_0 = 1000$, $I_0 = 400$; small seed: $E_0 = 100$, $I_0 = 40$), and estimate the shared parameter vector $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$ by jointly fitting the two epidemics; the joint log-likelihood

is the sum over both trajectories.

Table 5.3 compares single- and two-epidemic inference under correct specification. The confidence-interval width for ν shrinks from 1.842 (single) to 0.070 (two), a factor of 26, and coverage rises from 77.8% to 96.5%. A similar gain is seen for c^* (width $0.072 \rightarrow 0.008$, coverage $73.3\% \rightarrow 97.0\%$). As in Chapter 3, joint fitting resolves the wide-interval problem of single-epidemic inference when the family is correctly specified.

Table 5.3: Single vs. two-epidemic inference under correct specification (Lognormal-generated data, Lognormal fit). Median (SD), mean 95% CI width, and empirical coverage across 200 replicates. Adding a second epidemic substantially sharpens inference for ν and c^* , both of which are poorly identified from a single epidemic.

Parameter	Single epidemic		Two epidemics		Coverage (%)	
	Median (SD)	Width	Median (SD)	Width	Single	Two
$\mathcal{R}_0$	2.996 (0.034)	0.116	3.000 (0.010)	0.041	84.7	96.0
ν	1.367 (0.323)	1.842	1.414 (0.018)	0.070	77.8	96.5
t_0	14.986 (0.332)	1.341	15.012 (0.196)	0.693	94.9	92.0
c^*	0.297 (0.022)	0.072	0.300 (0.002)	0.008	73.3	97.0

In contrast, when the misspecified Gamma model is compelled to explain two Lognormal-generated epidemics (Table 5.4), the structural error can no longer be absorbed by a diffuse compensation ridge. The Gamma fit retains narrow intervals, but the estimates are biased. The median estimate of ν is 1.194, with mean 95% Wald limits (1.172, 1.218), compared with the data-generating value 1.414; the relative bias is -15.5% , and coverage for both $\mathcal{R}_0$ and ν is 0.0%. The parameters t_0 and c^* remain reasonably estimated, confirming that the mismatch is concentrated in the heterogeneity estimate.

The homogeneous fit fails on three of the four parameters in the joint setting. The median estimate of t_0 is 10.05 days, with mean 95% Wald limits (9.851, 10.163), compared with the data-generating value 15. The median estimate of c^* is 0.236, with mean 95% Wald limits (0.234, 0.238), while the median estimate of $\mathcal{R}_0$ is 3.010, with mean 95% Wald limits (2.997, 3.027). Coverage is 0.0% for t_0 and c^* and 45.9% for $\mathcal{R}_0$. Joint fitting therefore plays a dual role: it sharpens inference under correct specification and exposes misspecification through coverage collapse. Figure 5.8 shows the corresponding MLE densities.

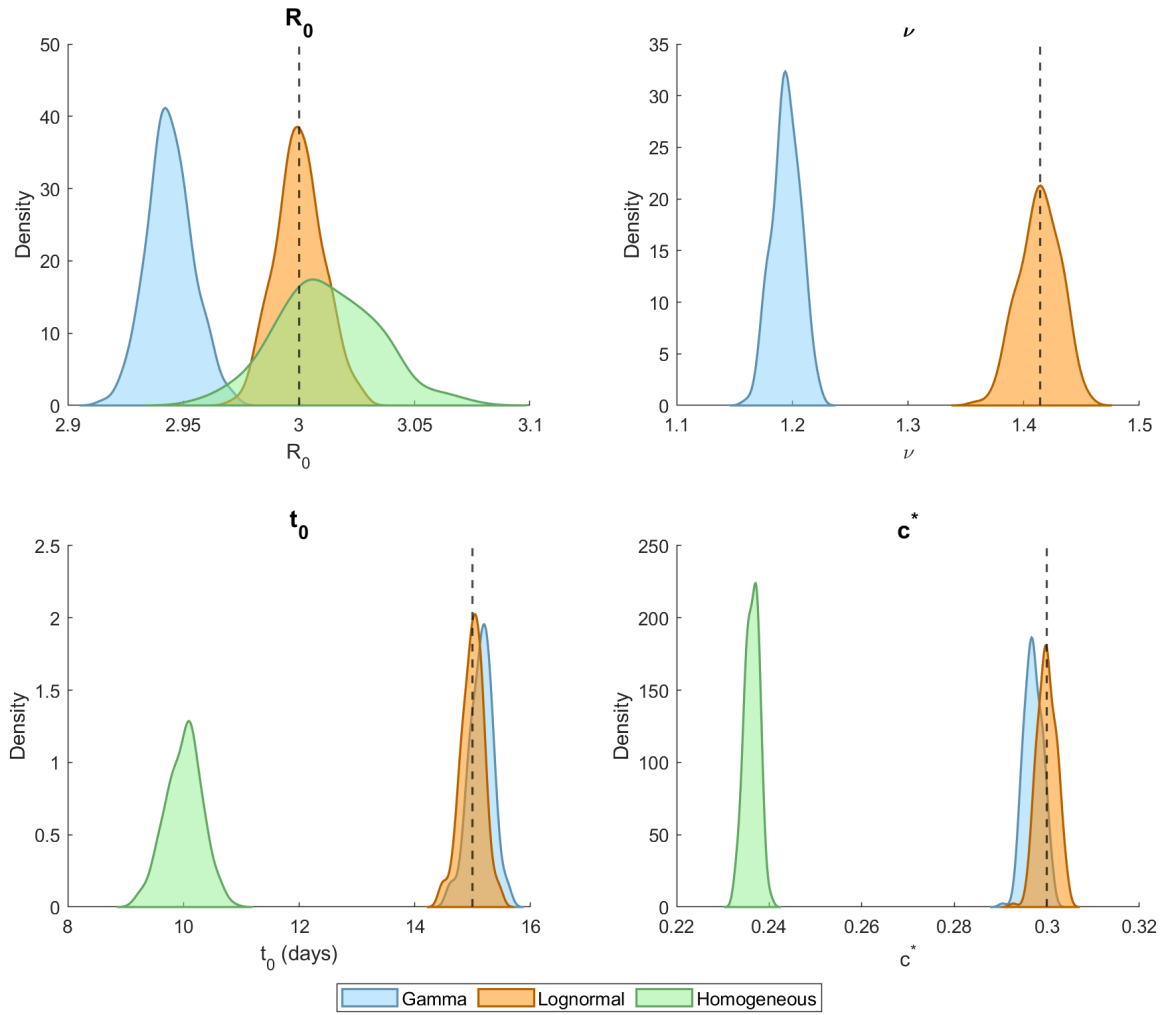


Figure 5.8: Distributions of two-epidemic MLE estimates for Lognormal-generated data ($\nu = 1.414$). Blue: Gamma (misspecified). Orange: Lognormal (correct). Green: homogeneous (reference). Dashed vertical lines indicate the data-generating values. The misspecified Gamma fit collapses to a sharp peak at a median ν estimate of approximately 1.19, with mean 95% Wald limits (1.172, 1.218), well separated from the correct Lognormal density. The homogeneous fit fails on t_0 and c^* .

Table 5.4: Two-epidemic inference for Lognormal-generated data, fitted with Lognormal, Gamma and homogeneous models ($\nu = 1.414$). Joint fit across large-seed and small-seed epidemics. Both heterogeneous fits produced valid interval summaries for all 200 replicates. For the homogeneous model, 193 optimisations converged and 181 yielded positive-definite Hessians and valid Wald intervals used in the interval summaries. Under correct specification, joint fitting recovers all parameters with near-nominal coverage; under misspecification, the Gamma fit concentrates on biased values for $\mathcal{R}_0$ and ν (0.0% coverage), while t_0 and c^* remain well estimated.

Model	Parameter	Median (SD)	CI width	Coverage (%)
Lognormal (correct)	$\mathcal{R}_0$	3.000 (0.010)	0.041	96.0
	ν	1.414 (0.018)	0.070	96.5
	t_0	15.012 (0.196)	0.693	92.0
	c^*	0.300 (0.002)	0.008	97.0
Gamma (misspecified)	$\mathcal{R}_0$	2.944 (0.010)	0.037	0.0
	ν	1.194 (0.012)	0.045	0.0
	t_0	15.154 (0.203)	0.689	82.0
	c^*	0.297 (0.002)	0.008	66.0
Homogeneous (reference)	$\mathcal{R}_0$	3.010 (0.022)	0.029	45.9
	ν	—	—	—
	t_0	10.051 (0.310)	0.312	0.0
	c^*	0.236 (0.002)	0.004	0.0

5.5.4 Fits and forecasts in joint inference (Lognormal baseline)

To translate the parameter biases into forecasts, we extend each two-epidemic simulation by lifting NPIs from day 100 onwards and projecting to day 250, computing forecasts per replicate from its joint-fit MLEs. Figure 5.9 shows the resulting trajectories.

In the training window, all three fits track the observed trajectories closely, as expected when the fitting window is long relative to the number of free parameters. The forecast period reveals the structural consequences. The correctly specified Lognormal fit tracks the deterministic data-generating trajectory in both epidemics, with a narrow band throughout. The misspecified Gamma fit, forced by the joint constraint to a median ν estimate of 1.194, with mean 95% Wald limits (1.172, 1.218), projects a smaller and earlier peak. The homogeneous fit departs from the data-generating trajectory already during the training window and overshoots substantially in both epidemics.

The forecast errors differ sharply between the two epidemics. In the small-seed epidemic, all three fits behave much as in the single-epidemic case: the correct Lognormal fit has a median peak-height error of +0.3%, the misspecified Gamma fit -1.3% , and the homogeneous fit $+270.3\%$. In the large-seed epidemic, susceptible depletion during the training window nearly prevents a second wave, whose data-generating peak height is only $2.4 \text{ cases day}^{-1}$. Consequently, small biases in ν produce large *relative* errors because the denominator is very small. The correct Lognormal

fit gives +1.5%, the misspecified Gamma fit -71.9% , and the homogeneous fit $+25,881\%$. The last percentage is inflated by the small data-generating peak used as its denominator and should therefore be interpreted with that caveat. The consistent ordering—homogeneous error $>$ misspecified heterogeneous error $>$ correctly specified heterogeneous error—holds in both epidemics.

At $\nu = 1.414$, therefore, the mean forecast trajectories are not alarmingly different for the two heterogeneous fits. The Gamma model underpredicts the peak, but its error is small compared with the gross overprediction of the homogeneous model discussed in Chapters 2 and 3. In a real application with moderate heterogeneity, a Gamma susceptibility model may produce practically adequate forecasts even when the data-generating distribution is Lognormal, at the cost of a systematic downward bias in peak magnitude that grows with the data-generating level of heterogeneity. The two-epidemic design exposes this bias clearly in the parameter estimates (Table 5.4), providing a diagnostic that the single-epidemic design cannot.

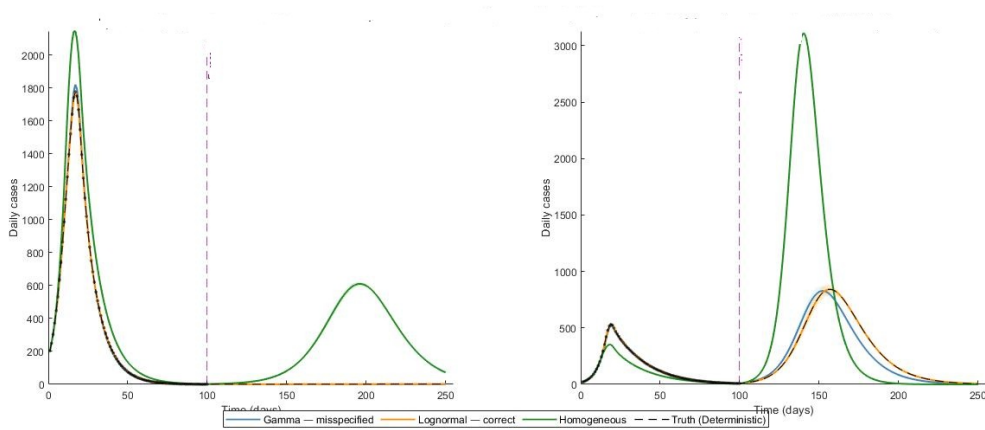


Figure 5.9: Two-epidemic trajectory fits and forecasts for Lognormal-generated data ($\nu = 1.414$, NPIs lifted at day 100 in each epidemic). The black dashed line is the deterministic data-generating trajectory; shaded bands are 95% confidence envelopes across replicated fits. The correctly specified Lognormal fit (orange) tracks both waves accurately. The misspecified Gamma fit (blue) tracks the data-generating trajectory closely within the 100-day training window but undershoots it in the forecast period. The homogeneous fit (green) departs from the data-generating trajectory during training and overshoots substantially in both epidemics, with the large-seed epidemic showing the largest relative error ($+25,881\%$ peak height compared with $+270\%$ in Epidemic 2).

5.5.5 Asymmetry across the two baselines

The complementary experiments in which the baseline distribution is Gamma (Appendix F) show the reverse directional signature on ν , but with larger biases. In two-epidemic inference at $\nu = 1.414$, fitting a Lognormal model to Gamma-generated data pushes the median estimate of ν upward to 1.781, with mean 95% Wald limits (1.768, 1.794) and relative bias $+25.9\%$ (Table F.3),

whereas the reverse analysis reported above gives a median estimate of 1.194, with mean 95% Wald limits (1.172, 1.218) and relative bias -15.5% . The Lognormal’s heavier tail implies greater susceptibility depletion than Gamma-generated data support, so the optimiser inflates ν to compensate; the magnitude asymmetry is a factor of $|+25.9\%|/|-15.5\%| \approx 1.7$.

The same asymmetry appears in the forecasts. Fitting the heavier-tailed Lognormal model to Gamma-generated data *overshoots* the second-wave peak, and does so more strongly than the lighter-tailed Gamma model *undershoots* when the roles are reversed: at $\nu = 1.414$ the Lognormal fit overshoots the small-seed epidemic peak by $+5.7\%$, compared with the -1.3% undershoot in the converse case (Figure F.4). Because misspecification is more costly in the heavier-tailed direction, it appears preferable to adopt the Gamma family as the default when the susceptibility distribution is unknown—a recommendation reinforced by the eigen-distribution property [43] and the computational tractability of the Gamma reduction [44, 52].

5.6 Chapter Summary, Discussion and Conclusion

This Chapter set out to determine whether the *shape* of the susceptibility distribution—specifically the difference between Gamma and Lognormal families—matters for epidemic forecasting when the first two moments are fixed (mean = 1 and a given coefficient of variation ν).

5.6.1 Discussion and Summary

Trajectory comparison (Section 5.3). When heterogeneity is low ($\nu \leq 0.5$), the epidemic curves from the Gamma and Lognormal models are nearly indistinguishable. As ν increases, a clear and systematic divergence emerges: the Lognormal model produces a higher peak, a later peak, and a larger final size than the matched Gamma model (Figure 5.3). At $\nu = 2.0$ with a fine discretisation ($k = 100$), the Lognormal peak incidence is approximately 31% higher and the final size approximately 49% larger than for the Gamma (Figure 5.4). The direction is always the same—Lognormal pessimistic, Gamma optimistic—driven by the heavier right tail of the Lognormal, which sustains transmission chains longer. Resolution k matters for cross-family comparisons: within-family computation is adequate at $k = 20$ (Chapter 4), but cross-family comparison requires $k \geq 40$; we used $k = 100$ throughout the inference experiments so that the tail signal is not compressed by coarse binning.

Single-epidemic inference (Section 5.5.1). for the Lognormal baseline, a correctly specified Lognormal fit recovers $\mathcal{R}_0$ and t_0 well but remains weakly informative for ν and c^* , consistent

with Chapters 2 and 3. The misspecified Gamma fit absorbs the structural error almost entirely through ν , deflating its median from 1.367 (correct) to 1.231 (relative bias -13.0%) while leaving the other parameters essentially unbiased (Table 5.2). Because the single-epidemic likelihood surface is nearly flat in the ν - c^* direction, the misspecification stays hidden: both heterogeneous fits produce similar in-sample fits and largely overlapping forecast bands (Figure 5.7). The homogeneous reference, however, is unambiguously wrong even from a single epidemic—collapsing c^* to 0.245 with 0% coverage and overshooting the forecast peak by $+216\%$, about 33 times the misspecified-Gamma error.

Single-epidemic inference for the Gamma baseline (Appendix F). The complementary Gamma-baseline experiment confirms that the main message does not depend on choosing the Lognormal baseline. Correct specification again recovers $\mathcal{R}_0$ and t_0 but is weakly informative for ν and c^* (Table F.1); the misspecified Lognormal fit now biases ν in the *opposite* direction, inflating its median to 1.607 (relative bias $+13.7\%$; Figure F.1). The forecast check (Figure F.2) shows the misspecified Lognormal fit matching the training window but projecting a higher, later second wave after NPI removal.

Two-epidemic inference (Section 5.5.3). Adding a second epidemic changes the picture substantially. With correct specification the confidence-interval width for ν shrinks from 1.842 to 0.070 and coverage rises from 77.8% to 96.5% (Table 5.3), confirming the identifiability gains of Chapter 2. Under misspecification the same sharpening forces the Gamma model into a tightly biased corner. The median estimate of ν is 1.194, with mean 95% Wald limits (1.172, 1.218) and relative bias -15.5% , while the median estimate of $\mathcal{R}_0$ is 2.944, with mean 95% Wald limits (2.926, 2.963). Coverage is 0.0% for both parameters (Table 5.4), while c^* remains essentially unbiased. The homogeneous reference fails on three of four parameters. The forecast consequence is visible but moderate at $\nu = 1.414$: the Gamma model undershoots the second-wave peak, but by far less than the homogeneous model overshoots it (Figure 5.9).

Two-epidemic inference for the Gamma baseline (Appendix F.1.3). Joint fitting again sharpens identifiability and exposes the misspecification. The misspecified Lognormal fit produces a median estimate of $\nu = 1.781$, with mean 95% Wald limits (1.768, 1.794) and relative bias $+25.9\%$, and a median estimate of $\mathcal{R}_0 = 3.221$, with mean 95% Wald limits (3.201, 3.241) and relative bias $+7.4\%$. Coverage is 0.0% for both parameters (Table F.3). The forecast consequence is more severe than in the Lognormal-baseline case: the Lognormal model overshoots

the second-wave peak with confidence bands that separate clearly from the data-generating peak (Figure F.4). This asymmetry–heavier-tailed misspecification producing a more alarming forecast error than lighter-tailed misspecification—is a recurring theme of the Chapter (Section 5.5.5).

Gamma as the benchmark. The Gamma distribution has served as our reference family for susceptibility, for three reasons. It is shape-preserving under epidemic selection, remaining Gamma as high-susceptibility individuals are depleted, so it is the natural benchmark for heterogeneity-induced braking (Chapter 1, Section 1.6.4). It has been applied successfully [4, 14, 44]. And, by the present results, it incurs the smaller forecast error under misspecification, since a lighter-tailed fit to heavier-tailed data undershoots less than the converse overshoots.

Limitations of distributional scope. This Chapter restricted the numerical experiments to the Gamma–Lognormal comparison, although Section 5.1 discussed the theoretical relevance of other families (Beta, Inverse Gaussian, Positive Stable). The Lognormal provides the most direct test of the heavy-tail hypothesis against the Gamma benchmark, and the distribution-agnostic discretisation of Chapter 4 was developed with the Lognormal as its worked example.

5.6.2 Conclusion

The main conclusion of this Chapter is that the choice of susceptibility distribution is not innocuous once heterogeneity is moderate to high. Even when Gamma and Lognormal models are matched on mean and coefficient of variation, the distributional shape produces differences in deterministic trajectories, systematic biases in parameter estimates, and directional errors in forecasts.

For single-epidemic data the practical impact is limited: the ν - c^* compensation ridge lets the misspecified model absorb the structural difference through a moderate shift in the estimated value of ν , and the forecast error stays within the wide uncertainty band that characterises single-epidemic inference. For two-epidemic data the impact is sharper: joint fitting removes the compensation flexibility and forces the misspecified model to commit to confidently incorrect parameter values, with systematic (and in the heavier-tailed direction, substantial) forecast errors. Across every scenario, however, the forecast cost of choosing the wrong heterogeneous family remains far smaller than the cost of ignoring heterogeneity altogether: the homogeneous reference overshoots peak incidence by hundreds of percent where the misspecified heterogeneous fit stays within a few percent to a few tens of percent.

Chapter 6

Model-Based Assessment of Heterogeneity in the UK COVID-19 Epidemic

6.1 Introduction

Chapters 2 to 5 demonstrated, under controlled simulation settings, that simultaneous inference of unobserved individual variation in susceptibility and the impact of non-pharmaceutical intervention (NPIs) presents a challenge in a single epidemic series. This can be alleviated considering two epidemics with similar heterogeneity but where NPIs are introduced at different times. This chapter moves to model-based inference from real data by fitting the heterogeneous SEIR models used in the foregoing Chapters to first wave of COVID-19 mortality in England and Scotland (January-June 2020, see Figure 6.1). Unlike the synthetic analyses, accurate daily incidences of infections were not readily available. Mortality, while more reliably recorded than cases early in the pandemic, is a delayed reflection of infections. Consequently, the inferential link from transmission to data requires an explicit observation layer that maps latent infections to observed deaths via an infection-to-death delay distribution and an infection fatality ratio (IFR). Wood et al [14] emphasise that deaths are naturally modelled through such a lagged convolution/deconvolution representation, and that negative binomial (or Poisson) observation noise can be used for daily death counts [14].

The approach in this chapter uses joint inference via *Bayesian Hierarchical Modelling* (see section A.3, Appendix A for basic concepts) to simultaneously fit the epidemic trajectories for England and Scotland. This approach connects directly to the identifiability lessons of Chapter

2 and 3 in particular. This chapter provides a platform to evaluate the drivers of curve flattening with three alternative specifications for the forms of NPIs. This also permits rigorous assessment of how sensitive inference of heterogeneity and intervention strength is to assumptions about the shape of contact reduction.

We also report a *comparison fit* similar to the additional results of Chapter 3 (Higher dimensional spaces in Bayesian inference) using uniform priors reported in Appendix D. This approach will be compared to Gomes et al [4] in section 6.8.4.

6.2 Data alignment

We analyse daily COVID-19 death counts for England and Scotland over the first-wave window (January–June 2020; see Figure 6.1). To ensure that country-specific intervention timing parameters can be interpreted on the exact same calendar scale, the two time series are aligned to a common temporal origin. Specifically, the daily index is anchored such that day 1 corresponds to January 31, 2020, aligning with the first recorded date in the joint dataset.

6.3 Transmission model

The reduced heterogeneous-susceptibility SEIR model was introduced in Chapter 1 and its derivation is given in Appendix A. We restate the operational form used in the mortality model. Fixed epidemiological inputs are the same as in Chapters 2–3; see Table G.1 in Appendix G.

6.3.1 Reduced SEIR with heterogeneous susceptibility

Let $S(t), E(t), I(t), R(t)$ denote the numbers of susceptible, exposed, infectious, and removed individuals in a population of size N . As in the earlier Chapters, heterogeneity in susceptibility is summarised by the coefficient of variation ν , with the homogeneous SEIR model recovered at $\nu = 0$.

On the count scale, the transmission dynamics are written as

$$\frac{dS}{dt} = -\lambda(t)S(t), \tag{6.1}$$

$$\frac{dE}{dt} = \lambda(t)S(t) - \delta E(t), \tag{6.2}$$

$$\frac{dI}{dt} = \delta E(t) - \gamma I(t), \tag{6.3}$$

$$\frac{dR}{dt} = \gamma I(t), \tag{6.4}$$

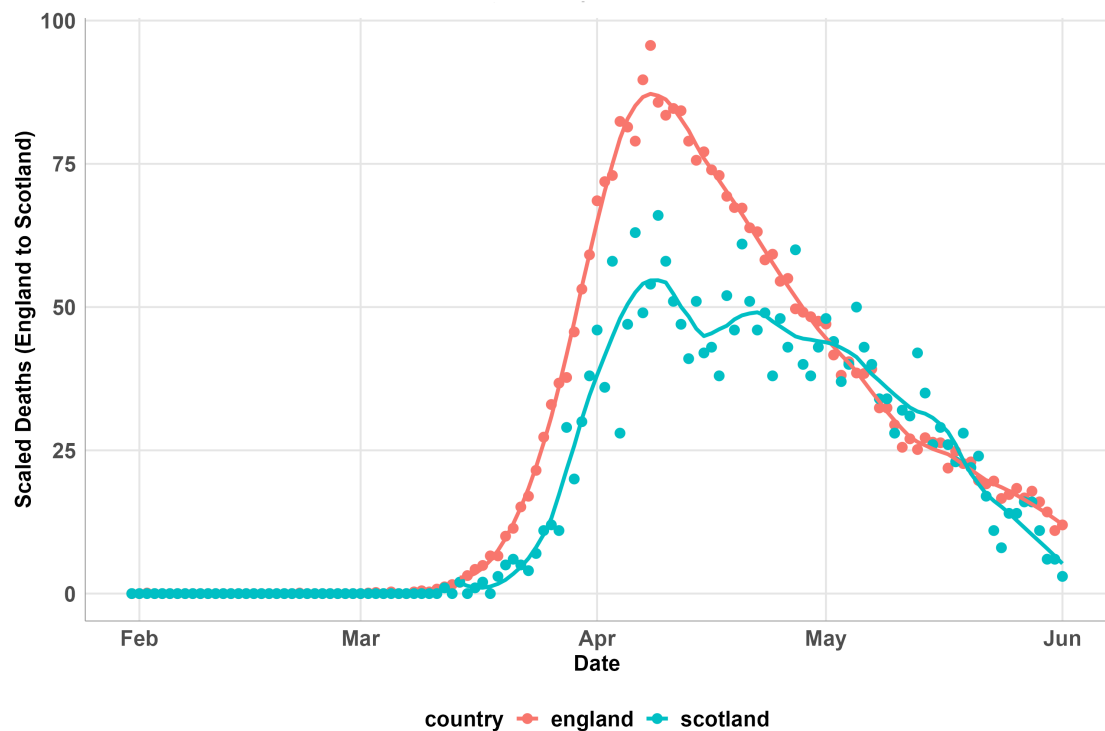


Figure 6.1: Daily reported COVID-19 deaths for England and Scotland over the first-wave analysis window (January–June 2020), aligned to a common calendar date. Points indicate daily counts, while solid curves represent smoothed trends to highlight the timing of the epidemic rise and peak. The steep rise in mortality begins earlier in England, leading to a higher, earlier peak and a faster decline, although the initial exponential growth slopes appear similar between the two nations. These distinct epidemic trajectories provide the necessary contrast for the two-epidemic inference to reduce the compensation between NPIs and population heterogeneity in susceptibility.

where the force of infection is

$$\lambda(t) = \beta(t) \frac{\rho E(t) + I(t)}{N} \left(\frac{S(t)}{N} \right)^{\nu^2}. \quad (6.5)$$

Equivalently,

$$\frac{dS}{dt} = -\beta(t) (\rho E(t) + I(t)) \left(\frac{S(t)}{N} \right)^{1+\nu^2}.$$

Thus, relative to the homogeneous SEIR model, the additional power term captures the progressive depletion of highly susceptible individuals.

The basic reproduction number is

$$\mathcal{R}_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right), \quad (6.6)$$

so that

$$\beta(t) = \frac{\mathcal{R}_0 c(t)}{\rho/\delta + 1/\gamma}. \quad (6.7)$$

The fixed quantities δ, γ, ρ are listed in Table G.1.

6.4 NPI parameterisations

We will use three parameterisations of the NPIs in this chapter:

1. Piecewise linear;
2. Logistic decay; and
3. Stringency index.

We have already used the first two in the preceding Chapters and will rewrite them for clarity here.

6.4.1 Piecewise-linear contact reduction

The baseline parametric specification is

$$c_{\text{lin}}(t) = \begin{cases} 1, & t \leq t_0, \\ 1 - (1 - c^*) \frac{t - t_0}{t_1 - t_0}, & t_0 < t \leq t_1, \\ c^*, & t > t_1, \end{cases} \quad (6.8)$$

where t_0 is the onset of behavioural change (estimated), t_1 is the start of official lockdown by Government which marks the beginning of maximal contact reduction, and $c^* \in (0, 1)$ is the sustained post-intervention contact level.

6.4.2 Logistic contact reduction

The second specification replaces the baseline piecewise-linear by a smooth logistic transition:

$$c_{\log}(t) = 1 - (1 - c^*) \logit^{-1}(k(t - t_m)) = 1 - \frac{1 - c^*}{1 + \exp\{-k(t - t_m)\}}, \quad (6.9)$$

where t_m is the midpoint of the transition and k controls its steepness. This parametrisation avoids the corners of the linear ramp and gives separate timing and steepness parameters for the decline in contact.

6.4.3 OxCGRT stringency-driven contact profile

As a third specification, we use the Oxford COVID-19 Government Response Tracker (OxCGRT) Stringency Index as an external time-varying policy signal [39, 100]. Let $SI_{j,t} \in [0, 100]$ denote the daily stringency value for country j , and rescale it to

$$s_{j,t} = \frac{SI_{j,t}}{100} \in [0, 1].$$

We then define the corresponding openness series by

$$u_{j,t} = 1 - s_{j,t},$$

and map it into the model contact modifier through

$$c_{j,t} = c_j^* + (1 - c_j^*) u_{j,t}, \quad (6.10)$$

where $c_j^* \in (0, 1)$ is the minimum contact level under maximal restrictions. This guarantees $c_{j,t} \in [c_j^*, 1]$ for all t . Throughout, smaller values of $c_j(t)$ indicate stronger contact reduction, while smaller values of c_j^* indicate a lower sustained contact level under maximal restrictions.

6.5 Observation model: from infections to deaths

6.5.1 Daily infections

Following Wood et al. [14], we model the infection-to-death delay using a Log-normal distribution and discretise it to daily probabilities for the convolution from infections to deaths. They report an infection-to-death duration model $\log(d) \sim \mathcal{N}(3.151, 0.469^2)$, constructed by combining an infection-to-onset meta-analysis with onset-to-death data, and use a negative binomial likelihood for daily deaths.

Let $\Delta N_{SE}(t)$ denote the number of new infections (new $S \rightarrow E$ transitions) occurring on day t . In the discrete-time implementation, $\Delta N_{SE}(t)$ is computed from the model trajectory as the daily incidence implied by (6.1).

6.5.2 Delay Convolution

Observed deaths are linked to infections through an infection-to-death delay distribution. Wood et al. [14] present a basic deconvolution representation for deaths in which expected deaths defined as $\mu(t)$ on day t_i are a lagged mixture of past infections with delay probabilities $g(l)$, and the resulting counts can be modelled using negative binomial (or Poisson) noise [14]. We adopt the equivalent forward (convolution) form:

$$\mu(t) = \psi \sum_{\ell=1}^L \Delta N_{SE}(t - \ell) g(\ell), \quad (6.11)$$

where $\psi \in (0, 1)$ is the infection fatality ratio, and $g(\ell)$ is the discretised infection-to-death delay probability mass function:

$$g(\ell) = \Pr(\ell - 1 < D \leq \ell) = F_D(\ell) - F_D(\ell - 1), \quad D \sim \text{Log-normal}(\mu_d, \sigma_d). \quad (6.12)$$

In the discrete-time implementation, $\Delta N_{SE}(t)$ is computed from the model trajectory as the daily incidence implied by (6.1). Figure 6.2 summarises the observation-model pipeline linking infections to observed deaths.

6.6 Bayesian Inference Framework

The observation model in Section 6.5 defines how the latent epidemic trajectory is linked to the reported daily deaths. Bayesian inference then combines this likelihood with prior distributions on the unknown parameters to obtain their posterior distribution.

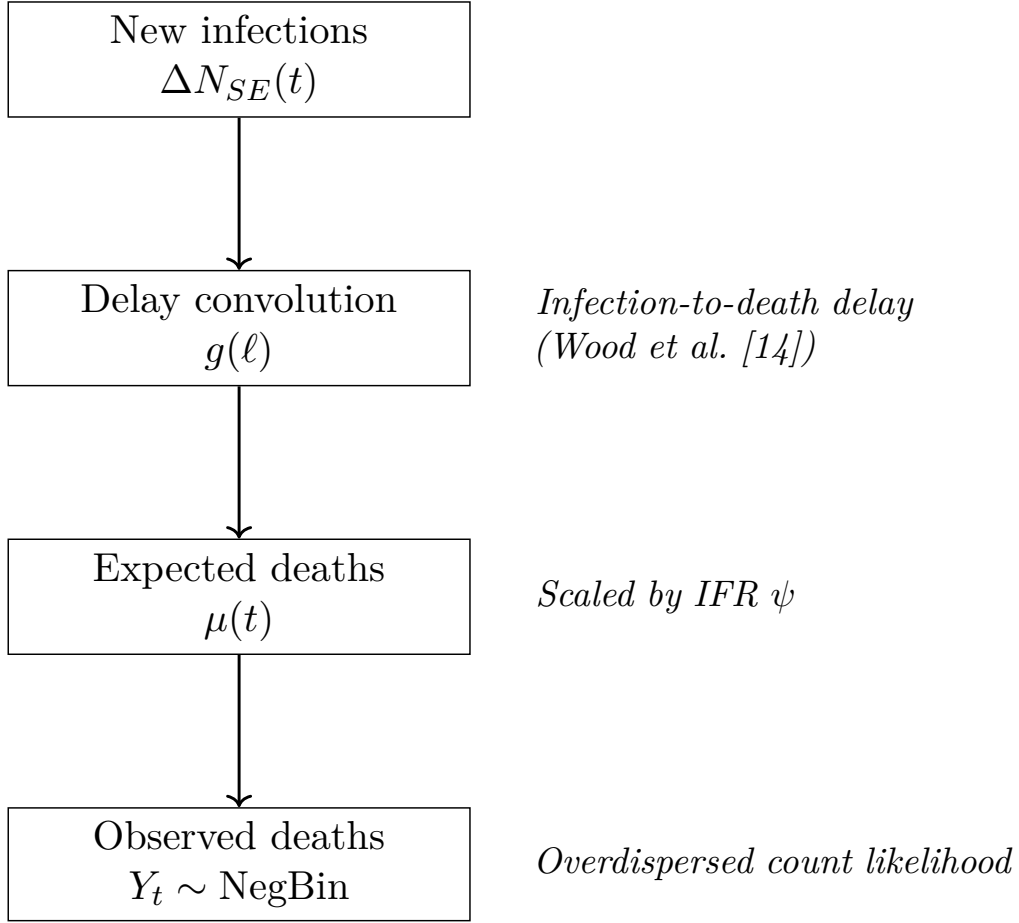


Figure 6.2: Observation-model pipeline linking latent infections to observed deaths. New infections $\Delta N_{SE}(t)$ are mapped to expected deaths through the infection-to-death delay distribution $g(\ell)$, scaled by the infection fatality ratio ψ , and then linked to observed daily deaths using a Negative Binomial likelihood.

6.6.1 Likelihood and posterior distribution

Let $Y_{j,t}$ denote the observed number of deaths in country j on day t , and let $\mu_j(t)$ denote the corresponding model-implied expected deaths from the delay convolution in equation (6.11). We model the daily death counts using a Negative Binomial distribution,

$$Y_{j,t} \sim \text{NegBin}(\mu_j(t), \phi_j),$$

where $\phi_j > 0$ is the country-specific dispersion parameter. With the mean-size parameterisation used here,

$$\text{Var}(Y_{j,t}) = \mu_j(t) + \frac{\mu_j(t)^2}{\phi_j}.$$

This form allows for overdispersion relative to a Poisson model, which is appropriate for daily mortality data.

Assuming conditional independence across countries and days given the parameter vector θ ,

the joint likelihood is

$$\mathcal{L}(\theta; \mathcal{Y}) = \prod_{j=1}^J \prod_{t=1}^T \Pr(Y_{j,t} \mid \mu_j(t \mid \theta), \phi_j), \quad (6.13)$$

with log-likelihood

$$\log \mathcal{L}(\theta; \mathcal{Y}) = \sum_{j=1}^J \sum_{t=1}^T \log \Pr(Y_{j,t} \mid \mu_j(t \mid \theta), \phi_j). \quad (6.14)$$

Bayesian inference targets the posterior distribution

$$p(\theta \mid \mathcal{Y}) \propto \mathcal{L}(\theta; \mathcal{Y}) p(\theta), \quad (6.15)$$

where $p(\theta)$ is the prior distribution and $p(\theta \mid \mathcal{Y})$ is the posterior distribution after observing the mortality data.

6.6.2 Parameter vector and hierarchical structure

The full parameter vector θ contains transmission parameters, NPI parameters, the shared susceptibility heterogeneity parameter ν , observation parameters, initial conditions, and the population-level hyperparameters used for partial pooling. To make the model specification explicit, it is useful to distinguish between the country-specific parameter vectors and the hyper-parameters that define the shared population distributions.

Let $m \in \{\text{lin}, \text{log}, \text{str}\}$ denote the linear, logistic, stringency index NPI specification respectively. For each country $j \in \{\text{EN}, \text{SC}\}$, we write the country-specific parameter vector as $\theta_j^{(m)}$.

For the **piecewise linear** specification,

$$\theta_j^{(\text{lin})} = (\mathcal{R}_{0,j}, c_j^*, w_j, \phi_j, I_{0,j}), \quad t_{0,j} = t_{1,j} - w_j,$$

together with the shared heterogeneity parameter ν . Here w_j is the duration of the transition, and $t_{0,j}$, $t_{1,j}$ as defined in equations (6.8).

For the **logistic** specification,

$$\theta_j^{(\text{log})} = (\mathcal{R}_{0,j}, c_j^*, t_{m,j}, k_j, \phi_j, I_{0,j}),$$

again together with the shared heterogeneity parameter ν . Here $t_{m,j}$ is the midpoint of contact reduction and k_j controls the steepness of the transition as in equation (6.9)

For the **stringency-driven** specification (equation (6.10)),

$$\theta_j^{(\text{str})} = (\mathcal{R}_{0,j}, c_j^*, \phi_j, I_{0,j}),$$

with ν again shared across the two countries. In this case, the time-varying contact profile is supplied externally by the OxCGRT-based openness series, so there are no additional timing or shape parameters within $c_j(t)$ beyond c_j^* . For all cases, $\mathcal{R}_{0,j}$ denotes the country-specific basic reproduction number, as defined in equation (6.6). The parameter ϕ_j is the country-specific Negative Binomial dispersion parameter in the likelihood (6.13), and $I_{0,j}$ is the initial infectious count used to seed the epidemic trajectory, with the initial exposed count linked to it through equation (6.17).

The corresponding full country-level parameter vector for each model can therefore be written as

$$\theta^{(m)} = (\theta_{\text{EN}}^{(m)}, \theta_{\text{SC}}^{(m)}, \nu), \quad m \in \{\text{lin}, \text{log}, \text{str}\}.$$

In the hierarchical models, however, these country-specific parameters are not estimated independently. Instead, selected components are allowed to vary around shared population-level distributions, so that fittings to England and Scotland can borrow strength from one another while still retaining country-level differences. For a pooled latent parameter ϑ_j , we use the non-centred parameterisation

$$z_{j,\vartheta} \sim \mathcal{N}(0, 1), \quad \vartheta_j = \mu_\vartheta + \sigma_\vartheta z_{j,\vartheta}, \quad (6.16)$$

where μ_ϑ is the population-level mean and σ_ϑ is the population-level standard deviation. This is equivalent to $\vartheta_j \sim \mathcal{N}(\mu_\vartheta, \sigma_\vartheta^2)$, but the non-centred form is often more efficient for Hamiltonian Monte Carlo [101].

The pooled parameters are defined on transformed scales. In particular

- **Piecewise Linear:** Pooled parameters $\vartheta_j \in \{\mathcal{R}_{0,j}, t_{0,j}, c_j^*\}$.
- **Logistic Function:** Pooled parameters $\vartheta_j \in \{\mathcal{R}_{0,j}, t_{m,j}, c_j^*\}$.
- **Stringency Index (OxCGRT):** Pooled parameters $\vartheta_j \in \{\mathcal{R}_{0,j}, c_j^*\}$.

The heterogeneity parameter ν is shared across both countries, while the Negative Binomial dispersion parameters ϕ_j and the initial infectious counts $I_{0,j}$ remain country-specific.

To summarise the hierarchical structure, if $\eta^{(m)}$ denotes the collection of population-level hyperparameters for model m , then the full hierarchical unknowns can be written schematically

as

$$\Theta^{(m)} = (\theta_{\text{EN}}^{(m)}, \theta_{\text{SC}}^{(m)}, \nu, \eta^{(m)}),$$

where $\eta^{(m)}$ contains the relevant population means and standard deviations for the pooled latent parameters under that NPI specification.

Prior specification and initial conditions

Weakly informative priors are assigned to the population-level hyperparameters in order to regularise the inference and keep the parameter search within plausible regions. Full prior details are reported in Appendix G, Table G.2.

For each country j , the initial infectious count $I_{0,j}$ is estimated, and the initial exposed count is linked to it through the relation (see Appndix A, subsection A.1.3 for derivation)

$$E_{0,j} = 2.5 I_{0,j}. \quad (6.17)$$

MCMC Convergence Diagnostics

All reported models were fitted using HMC with 4 chains and 2500 iterations per chain (1000 warmup).

6.6.3 Posterior predictive checks and fit metrics

To assess how well the fitted models reproduce the mortality data, we generate posterior predictive replicated deaths. For posterior draw $\theta^{(s)}$,

$$Y_{j,t}^{\text{rep},(s)} \sim \text{NegBin}(\mu_j(t | \theta^{(s)}), \phi_j^{(s)}). \quad (6.18)$$

These replicated trajectories are compared with the observed deaths using fitted trajectory plots, predictive intervals, and residual diagnostics.

For model-fit summaries, define the residual

$$r_{j,t} = Y_{j,t} - \hat{\mu}_{j,t}, \quad (6.19)$$

where $\hat{\mu}_{j,t}$ denotes the fitted posterior median death count for country j on day t . We report the

following summary measures:

$$\text{RMSE}_j = \sqrt{\frac{1}{T} \sum_{t=1}^T r_{j,t}^2}, \quad (6.20)$$

$$\text{MAE}_j = \frac{1}{T} \sum_{t=1}^T |r_{j,t}|, \quad (6.21)$$

$$\text{Bias}_j = \frac{1}{T} \sum_{t=1}^T r_{j,t}. \quad (6.22)$$

Here RMSE is the Root Mean Squared Error, MAE is the Mean Absolute Error, and Bias is the mean observed-minus-fitted error. Smaller RMSE and MAE indicate closer agreement between the fitted and observed mortality series, while Bias indicates whether the model tends to underpredict or overpredict on average.

6.7 Results

We now report the results of joint England-Scotland fitting described in the previous section. Table 6.1 summarises goodness-of-fit using the residual-based criteria defined in Section 6.6.3: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Bias.

6.7.1 Model Fit To Observed Mortality

Before examining the inferred transmission dynamics, we first evaluate how well each hierarchical specification fits the observed daily mortality data. Figures 6.3 and 6.4 display the posterior predictive trajectories for the reported deaths, alongside the implied forms of $c(t)$ for each NPI specification. Visually, all three parametrisations capture the first-wave rise and subsequent decline in both nations, with the peak timing particularly well reproduced for Scotland. Table 6.1 summarises fit using the residual-based criteria defined in equations (6.20)–(6.22): Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Bias.

To complement these aggregate summaries, Figure 6.5 plots residuals (observed minus fitted) over time. For England, the logistic specification reduces systematic residual structure around the peak compared with the piecewise linear and stringency-based models, consistent with its lower RMSE and MAE. However, the same figure also shows that all three NPI specifications retain some systematic lack of fit around the England mortality peak and during the post-lockdown decline: the fitted curves tend to under-predict deaths around the main peak and remain somewhat too high during the subsequent decline. The observation-layer sensitivity analysis in Section 6.7.5 examines this residual pattern more directly. The main point is that

Table 6.1: Goodness-of-fit metrics for the three hierarchical NPI models fitted to daily mortality data in England and Scotland. Metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Bias (mean observed minus fitted).

Country	Model	RMSE	MAE	Bias
England	Linear	48.1	30.4	1.34
England	Logistic	36.4	21.7	2.60
England	Stringency	47.4	31.5	-2.74
Scotland	Linear	5.91	3.63	-0.08
Scotland	Logistic	6.00	3.69	-0.12
Scotland	Stringency	5.75	3.66	0.26

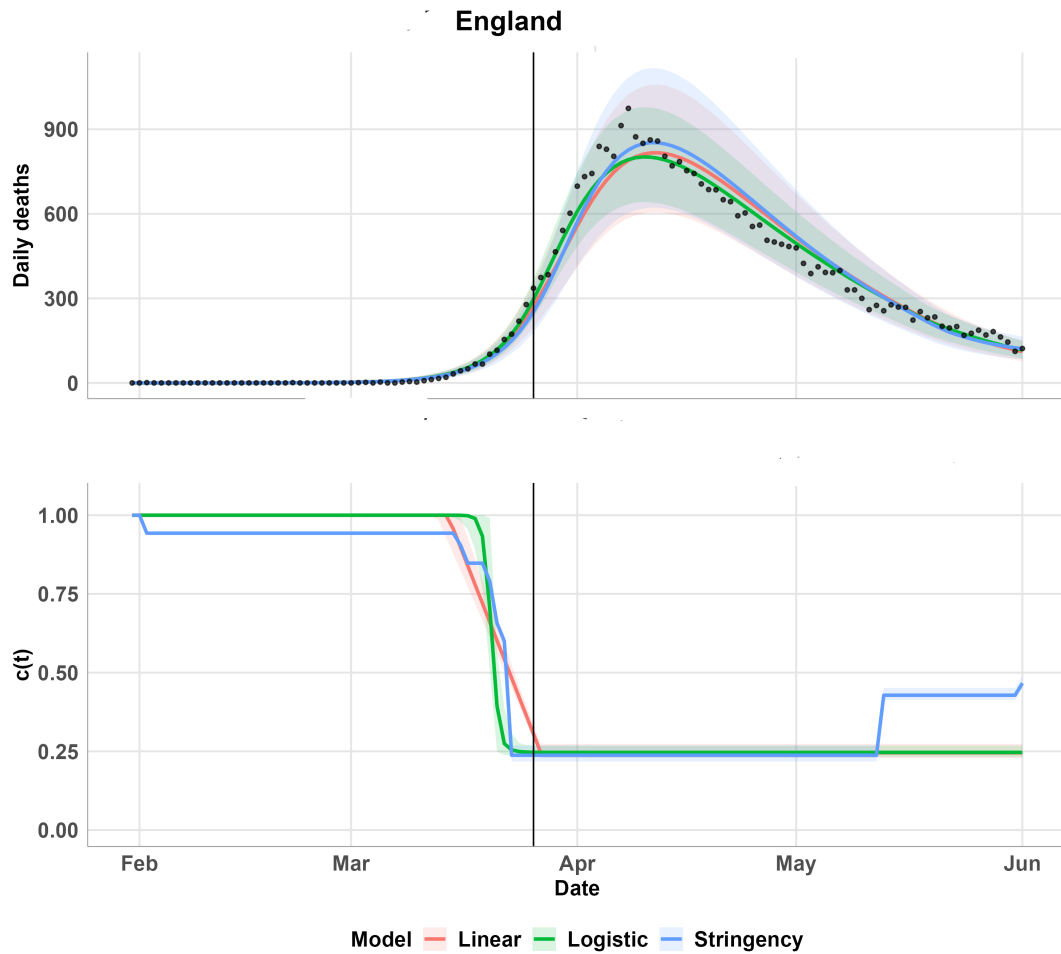


Figure 6.3: Fitted trajectories of England for the three hierarchical NPI specifications. The panels display the posterior medians (solid lines) and 95% pointwise credible intervals (shaded ribbons) for the piecewise linear (red), logistic (green), and empirical stringency (blue) models. **Top:** Model fit to observed daily COVID-19 mortality. All three functional forms reasonably capture the features of the mortality curve. **Bottom:** The inferred NPI $c(t)$ over time. The vertical black line denotes the official UK lockdown date (26 March 2020 for England). Despite differing onset times and transition speeds, all models converge to a consistent contact reduction ($c^* \approx 0.24$).

the broad infection-to-death delay used in the fitted mortality model can smooth and delay the death curve relative to the latent infection peak. The England residual pattern is therefore best

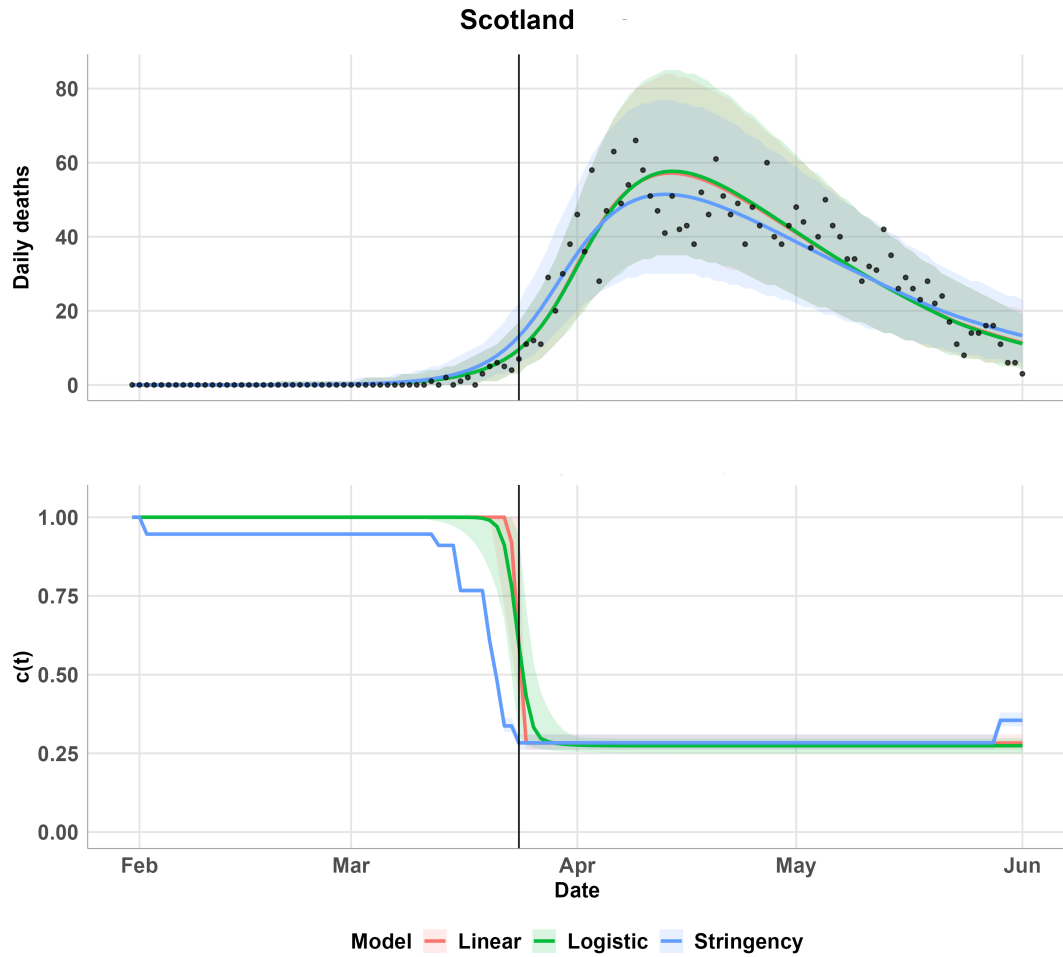


Figure 6.4: Fitted trajectories for Scotland using three hierarchical NPI specifications. Posterior medians (solid lines) and 95% credible intervals (shaded ribbons) are shown for the linear (red), logistic (green), and stringency (blue) models. **Top:** Model fit to observed daily mortality. The models accommodate the higher relative noise in the Scottish series while capturing the overall epidemic wave. **Bottom:** The inferred NPI $c(t)$ relative to the official lockdown date (vertical black line: 24 March 2020 for Scotland). Compared to England, the parametric models infer a distinctly later onset of behavioural change, followed by a much sharper drop to the sustained intervention level ($c^* \approx 0.28$).

interpreted as a limitation of the combined transmission–observation framework, rather than as evidence against any single NPI functional form alone. For Scotland, residual patterns are broadly comparable across models, in line with the near-identical fit metrics and the weaker systematic structure in the Scottish series.

6.7.2 Posterior Summaries Of Shared Population-level Parameters

We now compare population-level summaries across the three hierarchical NPI specifications: the basic reproduction number $\mathcal{R}_{0,\text{pop}}$, the minimum contact level c_{pop}^* , and the shared susceptibility heterogeneity parameter ν . Here $\mathcal{R}_{0,\text{pop}}$ and c_{pop}^* denote posterior-draw summaries obtained by averaging the corresponding country-specific quantities over England and Scotland within

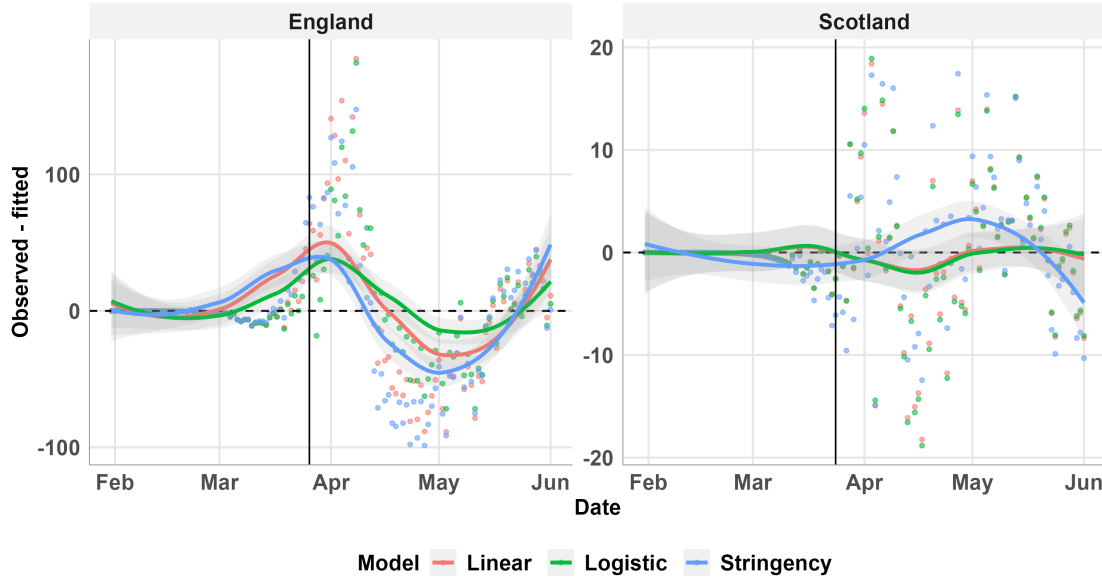


Figure 6.5: Residuals for the three joint hierarchical NPI models, shown separately for England and Scotland. Points are daily residuals; smooth curves summarise residual trends over time for each NPI specification. The vertical line denotes the official lockdown date. This plot complements Table 6.1 by revealing time-local misfit (e.g., around the peak) that may not be fully reflected in RMSE/MAE summaries.

each NPI model. Table 6.2 reports posterior medians, posterior SDs, and central 95% credible intervals.

Intervention and baseline transmissibility. Across all three NPI specifications, the inferred population-level contact floor is stable, with c_{pop}^* concentrated around 0.26–0.27 (Table 6.2 and Figure 6.6). This implies that contacts were reduced to roughly 26–27% of their pre-pandemic baseline during the strictest phase of control. Similarly, the population-level basic reproduction number $\mathcal{R}_{0,\text{pop}}$ is stable across NPI specifications, with posterior medians between approximately 2.91 and 3.06.

Heterogeneity and its sensitivity to the assumed NPI shape. The main quantity of interest in this analysis is the centred susceptibility heterogeneity parameter (ν). With the hierarchical *logistic* model, the posterior median is $\nu = 1.03$ (95% CrI 0.48–1.56). The matched *piecewise linear* model yields a slightly higher median of $\nu = 1.21$ (0.65–1.79). Also, the empirically driven *stringency* model infers the highest level of heterogeneity, with a median of $\nu = 1.39$ (0.77–1.98). While the credible intervals overlap, this upward shift reveals a moderate sensitivity of inferred heterogeneity to the assumed contact-transition shape.

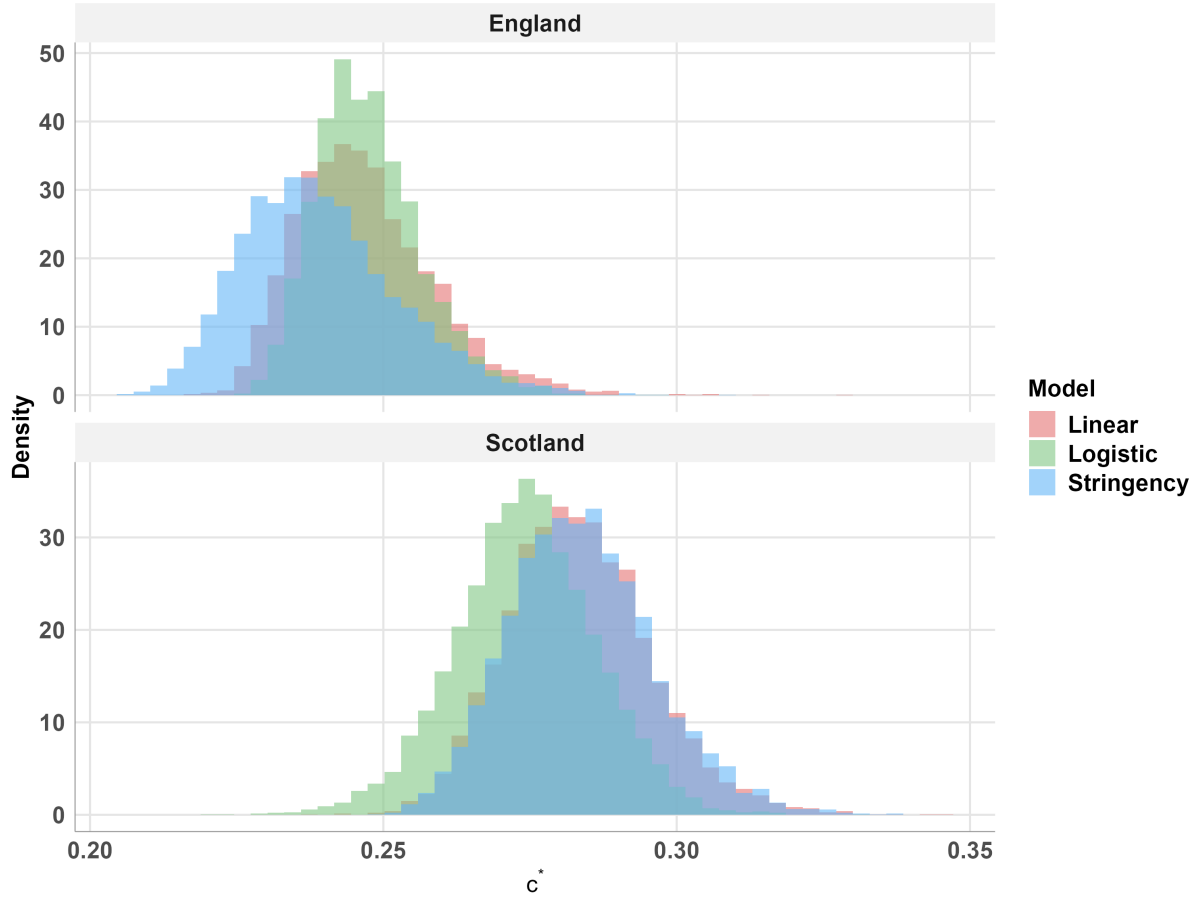


Figure 6.6: Posterior comparison of the minimum contact level c^* across the three joint hierarchical NPI specifications, shown separately for England and Scotland. Smaller values of c^* correspond to stronger maximal contact reduction. Within each country, the posterior distributions from the linear, logistic, and stringency models overlap substantially, indicating that the inferred lockdown contact level is fairly stable across NPI forms. This agrees with the country-level posterior summaries in Table 6.3 and the overlapping population-level summaries in Table 6.2.

Table 6.2: Joint hierarchical population-level posterior summaries for the three NPI specifications. Entries are posterior medians (Q50) with posterior standard deviations (SD) in parentheses and central 95% credible intervals in brackets. Here $\mathcal{R}_{0,\text{pop}}$ and c_{pop}^* denote per-draw averages of the corresponding country-specific posterior quantities over England and Scotland.

Parameter	Posterior summaries: Q50 (SD) [Q2.5, Q97.5]		
	Piecewise linear	Logistic	Stringency
$\mathcal{R}_{0,\text{pop}}$	2.91 (0.703) [2.25, 4.333]	2.93 (0.880) [2.22, 4.247]	3.06 (0.630) [2.41, 4.288]
c_{pop}^*	0.269 (0.0546) [0.182, 0.415]	0.264 (0.0506) [0.185, 0.403]	0.268 (0.0609) [0.182, 0.439]
ν	1.21 (0.298) [0.647, 1.79]	1.03 (0.273) [0.479, 1.56]	1.39 (0.308) [0.770, 1.98]

6.7.3 Country-specific posterior summaries

Table 6.3 reports country-specific posterior summaries for each NPI specification, including observation dispersion (ϕ_j), initial infections ($I_{0,j}$), and (where applicable) NPI timing/shape

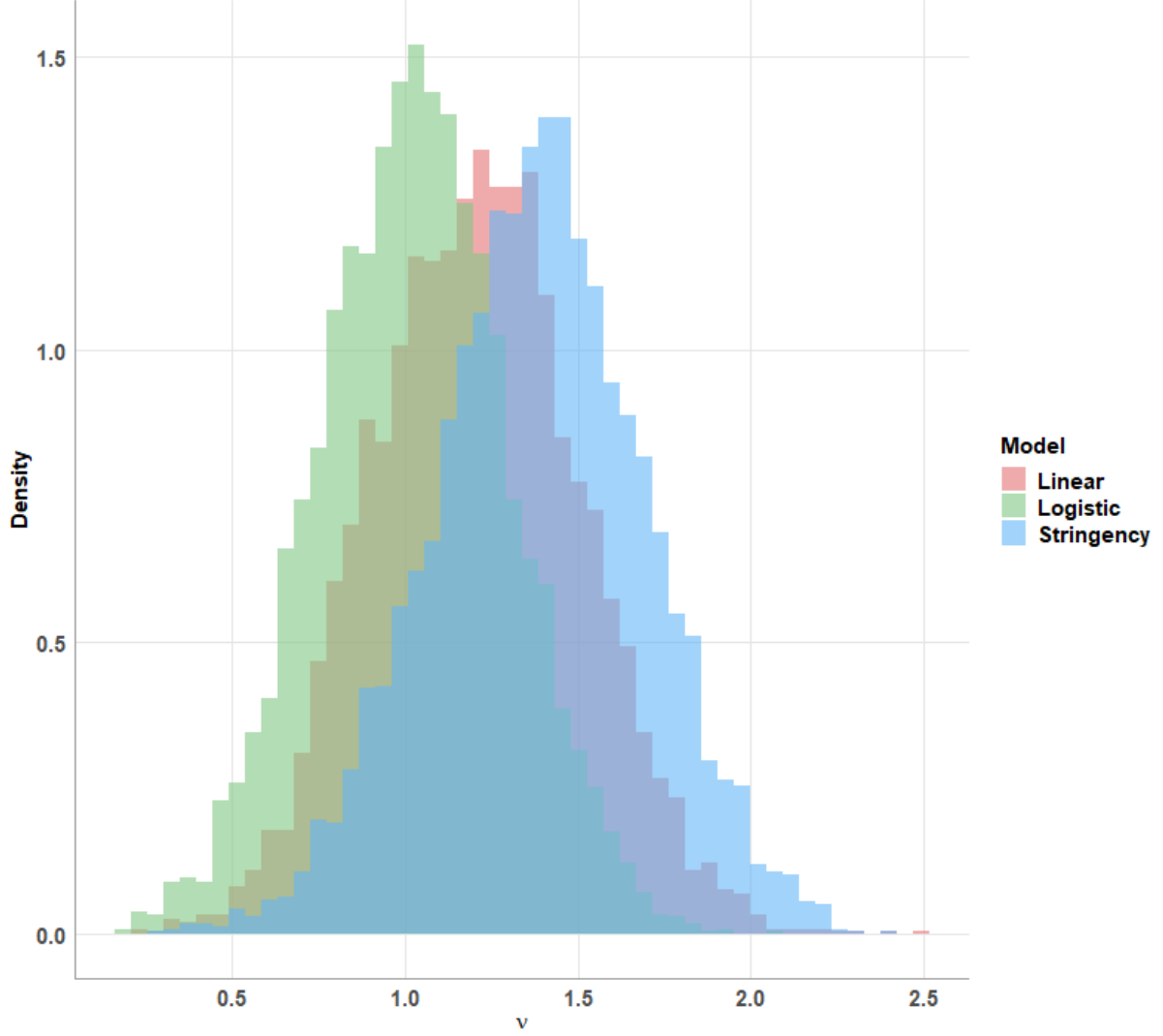


Figure 6.7: Posterior comparison of the shared susceptibility-heterogeneity parameter ν across the three joint hierarchical NPI specifications. The posterior distributions overlap, but their locations differ systematically across models: the logistic specification gives the lowest values, the piecewise linear model is intermediate, and the stringency-driven model gives the highest values. These are in line with the posterior summaries in Table 6.2, where ν shows more sensitivity to the assumed NPI form than either c^* or the pooled transmissibility summary.

parameters.

Country-level comparison for transmission and lockdown

At the country level, the basic reproduction number ($\mathcal{R}_{0,j}$) has a consistent difference in initial transmission dynamics between the two nations (see Figure 6.8). England’s baseline transmissibility for the piecewise linear model $\mathcal{R}_{0,EN} = 2.99$ (95% CrI 2.92-3.06) versus Scotland’s $\mathcal{R}_{0,SC} = 2.70$ (2.64-2.77). Results are similar for the stringency-driven model, both $\mathcal{R}_{0,EN}$ and $\mathcal{R}_{0,SC}$ shift upward (3.10 and 2.97, respectively), consistent with the population-level shift in ν under stringency. Despite these differences in $\mathcal{R}_0$, the inferred contact reduction remains

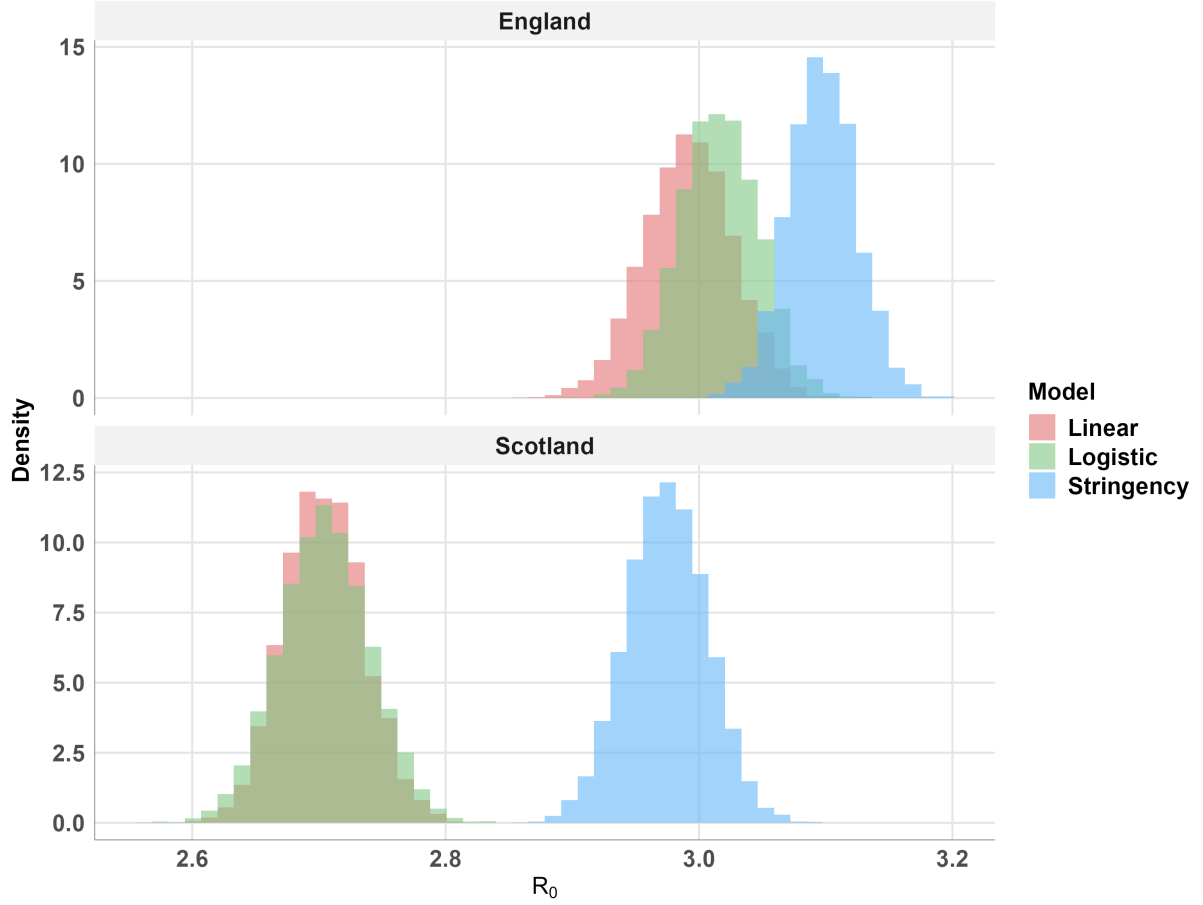


Figure 6.8: Posterior comparison of the country-specific basic reproduction number $\mathcal{R}_0$ across the three joint hierarchical NPI specifications, shown separately for England and Scotland. The linear and logistic models give similar posterior distributions within each country, whereas the stringency-driven model shifts $\mathcal{R}_0$ upward, especially for Scotland. The figure also shows that England and Scotland have distinct posterior locations, consistent with the country-level summaries in Table 6.3. At the same time, the pooled population-level summaries in Table 6.2 show that overall transmissibility changes only moderately across NPI forms.

highly robust at the country level. England’s c_{EN}^* concentrates near 0.24-0.25 across models, while Scotland’s c_{SC}^* concentrates near 0.27-0.28 (Table 6.3, Figures 6.6-6.8 to visualise the corresponding marginal posterior distributions). This stability is also evident in the $c(t)$ panels in Figures 6.3-6.4, where the three models converge to similar lockdown contact levels despite differing transition shapes as well as posterior plots in Figure 6.6.

Observation dispersion. Differences between the NPI shapes are more visible in the negative-binomial dispersion parameter (ϕ), particularly for England. With the *logistic* fit, England’s ϕ has a median of 97.8 (95%CrI 59.4-164), which indicate that relatively low overdispersion (a tighter fit to the Poisson distribution). However, under both the *piecewise linear* ($\phi = 53.0$) and *stringency* ($\phi = 48.0$) models, dispersion is higher. Scotland is much more consistent across

Table 6.3: Joint Hierarchical country-level posterior summaries across three NPI specifications. Entries are posterior median (Q50) with posterior standard deviation (SD) in parentheses and central 95% credible intervals in brackets. England rows are followed by Scotland rows. Parameters not present in a given NPI specification are shown as “-”. Time parameters are reported on the aligned day index used in the model. The centred heterogeneity parameter ν is reported separately at population level (Table 6.2).

Parameter	Posterior summaries (Q50 (SD) [Q2.5, Q97.5])		
	Piecewise linear	Logistic	Stringency
England			
$\mathcal{R}_{0,EN}$	2.99 (0.0350) [2.92, 3.06]	3.02 (0.0309) [2.96, 3.08]	3.10 (0.0270) [3.04, 3.15]
c_{EN}^*	0.245 (0.0116) [0.228, 0.274]	0.246 (0.0089) [0.233, 0.268]	0.238 (0.0132) [0.217, 0.268]
ϕ_{EN}	53.0 (12.3) [33.6, 81.9]	97.8 (26.8) [59.4, 164]	48.0 (10.2) [31.9, 72.1]
$I_{0,EN}$	30.1 (2.87) [25.1, 36.3]	28.3 (2.55) [23.8, 33.7]	32.9 (2.86) [27.9, 39.1]
$t_{0,EN}$	43.3 (0.828) [41.7, 44.9]	-	-
w_{EN} (days)	12.7 (0.828) [11.1, 14.3]	-	-
$t_{m,EN}$	-	49.2 (0.301) [48.7, 49.8]	-
k_{EN}	-	1.91 (3.00) [0.897, 9.67]	-
Scotland			
$\mathcal{R}_{0,SC}$	2.70 (0.0317) [2.64, 2.77]	2.70 (0.0361) [2.63, 2.78]	2.97 (0.0319) [2.91, 3.04]
c_{SC}^*	0.283 (0.0123) [0.262, 0.310]	0.274 (0.0121) [0.250, 0.297]	0.283 (0.0124) [0.262, 0.310]
ϕ_{SC}	33.2 (12.2) [18.6, 65.6]	31.9 (11.7) [17.3, 63.2]	29.3 (11.0) [15.7, 58.9]
$I_{0,SC}$	3.69 (0.365) [3.05, 4.48]	3.69 (0.364) [3.07, 4.48]	3.51 (0.344) [2.90, 4.25]
$t_{0,SC}$	51.8 (1.24) [49.0, 53.8]	-	-
w_{SC} (days)	2.25 (1.24) [0.242, 4.96]	-	-
$t_{m,SC}$	-	52.8 (0.785) [51.3, 54.4]	-
k_{SC}	-	1.16 (2.02) [0.390, 5.33]	-

ϕ_j is the Negative-Binomial dispersion for country j . $I_{0,j}$ is the inferred initial infectious count at the start of the fitted window. For the piecewise linear-model specification, $(t_{0,j}, w_j)$ describe onset and duration of contact reduction. w_j is estimated and $t_{0,j} := t_{1,j} - w_j$. For the logistic specification, $(t_{m,j}, k_j)$ describe midpoint and steepness. For the stringency-driven specification, the time-varying profile of $c(t)$ is driven by the observed OxCGRT series (no additional timing/shape parameters within $c(t)$ beyond c_j^*).

all three NPI shapes, with ϕ remaining stable around 29 to 33. Since ϕ controls the degree of observation-level noise in the Negative binomial likelihood, this pattern indicates that for England’s larger population, the smooth logistic transition leaves less residual variability for the observation layer to absorb compared to the piecewise or empirical NPI shapes.

Timing of behavioural change. While the *stringency* model relies on an external daily index, the other two parametric models infer the timing of contact reductions directly from the data, but encode it differently. For the logistic model, the midpoint parameter is tightly identified: $t_{m,EN} = 49.2$ (48.7–49.8), corresponding approximately to 19 March 2020 (18–20 March), and $t_{m,SC} = 52.8$ (51.3–54.4), corresponding approximately to 23 March 2020 (21–24 March). Thus, on the aligned day index, the inferred midpoint of contact reduction occurs later in Scotland than in England.

For the piecewise linear model, the same ordering appears in the onset time: $t_{0,EN} = 43.3$

(41.7–44.9), corresponding approximately to 13 March 2020 (12–14 March), compared with $t_{0,SC} = 51.8$ (49.0–53.8), corresponding approximately to 22 March 2020 (19–24 March). In contrast, the logistic steepness parameters are comparatively uncertain, for example $k_{EN} = 1.91$ with a wide 95% credible interval of 0.897–9.67. This indicates that the data constrain the *timing* of the transition more strongly than the *exact steepness* once the model is allowed to be smooth.

6.7.4 Model Comparison and Selection

We compared the three hierarchical NPI specifications using PSIS-LOO and WAIC, reported on the expected log pointwise predictive density (elpd) scale. Higher elpd indicates better expected out-of-sample predictive accuracy. Table 6.4 summarises the three-way comparison, with the best-performing model used as the reference ($\Delta\text{elpd} = 0$). Both criteria select the logistic transition model as the best predictive performer, followed by the piecewise linear model and then the stringency-driven model. The ordering is the same for PSIS-LOO and WAIC.

Table 6.4: PSIS-LOO and WAIC comparison across the three joint hierarchical NPI models. Entries are differences in elpd relative to the best model ($\Delta\text{elpd} = 0$), with standard errors of the differences in parentheses. Higher elpd indicates better expected out-of-sample predictive accuracy.

Model	PSIS-LOO	WAIC
Logistic	0.0 (0.0)	0.0 (0.0)
Piecewise linear	−18.5 (2.9)	−18.5 (2.9)
Stringency	−31.0 (5.9)	−31.0 (5.9)

The difference between the logistic and piecewise linear models is substantial: the piecewise linear model is lower by $\Delta\text{elpd} = -18.5$ with standard error 2.9. The stringency-driven model performs worse still, with $\Delta\text{elpd} = -31.0$ relative to the logistic model. The pairwise comparison in Appendix G also shows that the stringency-driven model is lower than the piecewise linear model ($\Delta\text{elpd} = -12.5$, s.e. = 4.6), confirming the ranking logistic, piecewise linear, stringency-driven.

This ordering agrees with the residual diagnostics in Figure 6.5, where the logistic model shows the least systematic residual structure in England. Detailed discussions and full pairwise comparisons and Pareto- k diagnostics are reported in Appendix G, Section G.3.

6.7.5 Observation-layer sensitivity: infection-to-death delay

The residual diagnostics in Figure 6.5 show that the remaining lack of fit in England is not fully explained by the choice of NPI functional form. All three NPI specifications show some discrepancy

around the mortality peak and during the post-lockdown decline. Since the observations are deaths rather than infections, this discrepancy may also arise from the observation layer linking latent infections to reported mortality.

We therefore examined the sensitivity of the fitted logistic model to the assumed infection-to-death delay distribution. The main results are summarised here because they affect the interpretation of the England residual pattern and the inferred balance between heterogeneity and contact reduction. Full posterior summaries and supporting details for the narrower-delay refit are reported in Appendix G, Section G.1.

To examine this, we carried out two sensitivity checks using the hierarchical logistic model (Table 6.3). The logistic specification is used here because it gives the best overall fit for England among the three hierarchical NPI forms in Table 6.1 as well as model selection in Section 6.7.4. First, we held the posterior median infection incidence fixed and remapped it to deaths using alternative infection-to-death delay assumptions. Second, we refitted the hierarchical logistic model using a narrower infection-to-death delay, $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$, instead of the baseline delay used in the main analysis with standard deviation of 0.469.

For the remapping exercise, let $\Delta N_{SE}(t)$ denote the posterior median latent infection incidence from the fitted logistic model. Expected deaths are then recomputed as

$$\mu(t) = \psi \sum_{\ell=1}^L \Delta N_{SE}(t - \ell) g(\ell), \quad (6.23)$$

where $\psi = 0.009$ is kept fixed and only the delay probabilities $g(\ell)$ are varied. For England, we compare the original fitted logistic death mean, a fixed-lag mapping with lag 12 days, and a narrower distributed delay with central 95% range approximately 15–40 days. The delay quantiles for the distributed mappings are shown in Table 6.5.

Table 6.5: Delay quantiles for the two distributed-delay mappings used in the logistic remapping sensitivity analysis. The fixed-lag comparator is a point mass at 12 days and therefore does not have comparable quantiles.

Distribution	Q2.5	Q50	Q97.5
Wood baseline [14]	9.32	23.4	58.6
Narrower delay	15.0	24.5	40.0

Figure 6.9 shows that the same fitted (of section 6.7) infection incidence can imply noticeably different death curves using different delay assumptions. The broad Wood’s [14] delay produces a smoother and later death curve. The fixed-lag mapping produces an earlier and sharper peak, while the narrower distributed delay gives an intermediate shape.

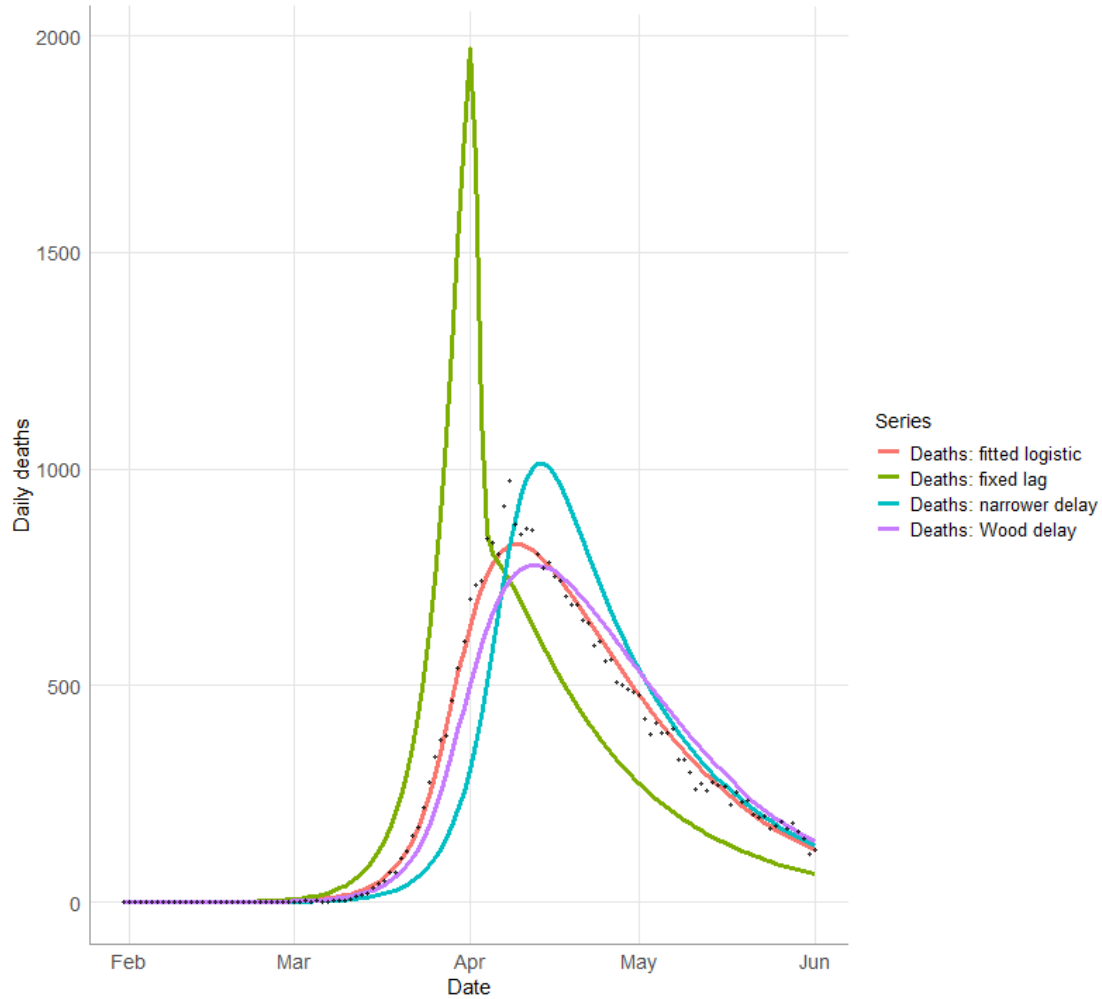


Figure 6.9: Expected daily deaths in England obtained by remapping the same posterior median infection incidence from the hierarchical logistic fit using alternative infection-to-death assumptions. Black points show observed daily deaths. Curves compare the original fitted logistic death mean, a fixed-lag mapping, and a narrower distributed delay. The fixed-lag mapping gives an earlier and sharper peak, while the narrower distributed delay gives an intermediate shape. This shows that part of the remaining England misfit can arise from the observation layer, not only from the assumed NPI form.

Table 6.6: England delay-remapping sensitivity using the posterior median infection incidence from the hierarchical logistic fit. Peak index is counted on the aligned day scale used throughout Chapter 6.

Method	Peak index	Peak value	Total
Logistic fit incidence, posterior median	50	218840	3935831
Deaths from logistic incidence: fixed lag	62	1970	34885
Deaths from logistic incidence: narrower delay	75	1013	33541
Fitted $\mu(t)$ (column 2, Table 6.3)	70	826	33762

The numerical summaries in Table 6.6 support the visual comparison. The posterior median infection incidence peaks on day 50, whereas the original fitted logistic death mean peaks on day 70 with peak magnitude about 826. With a fixed 12-day lag, the implied death peak moves earlier

to day 62 and becomes much sharper, with peak magnitude about 1970. Under the narrower distributed delay, the peak height is intermediate, about 1013, and occurs on day 75. Thus, the assumed delay distribution can materially affect the fitted mortality shape even when the latent infection incidence is unchanged.

The second sensitivity check refits the hierarchical logistic model using the narrower delay $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$. This refit asks whether reducing the smoothing induced by the observation layer improves the mortality fit when the transmission model is re-estimated rather than merely remapped. Figures 6.10 and 6.11 show the fitted death curves and residual diagnostics.

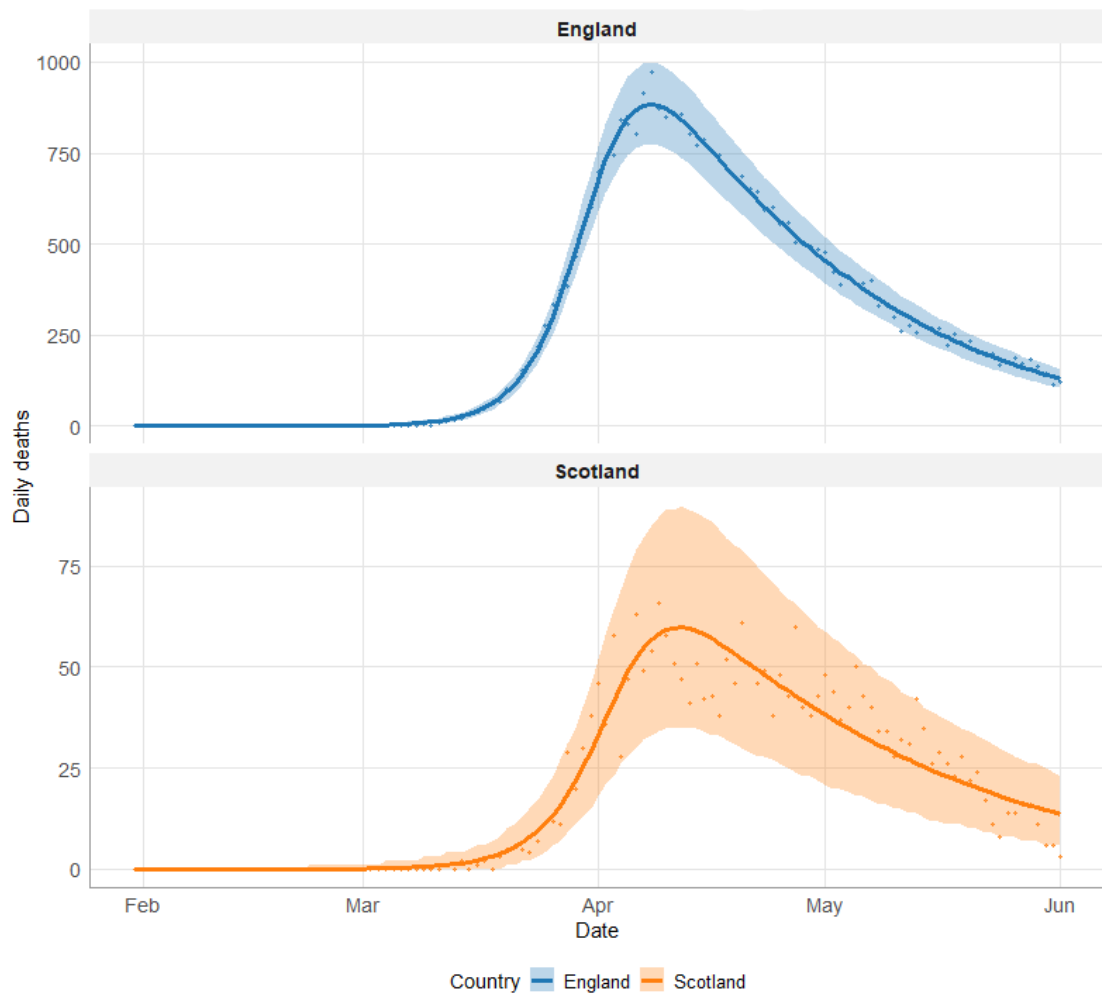


Figure 6.10: Fitted daily deaths for the hierarchical logistic model refitted with infection-to-death delay $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$, shown for England and Scotland. Solid lines are posterior medians and shaded bands are 95% pointwise credible intervals; points are observed daily deaths. Compared with the baseline logistic fit, the narrower delay gives a closer fit around the England peak and post-lockdown decline, while the fitted trajectory for Scotland changes less.

The narrower-delay refit improves the visual alignment around the England peak and post-lockdown decline, but it also changes the inferred parameters. Table 6.7 summarises the main shift. At the population level, the pooled baseline transmissibility remains broadly similar, but

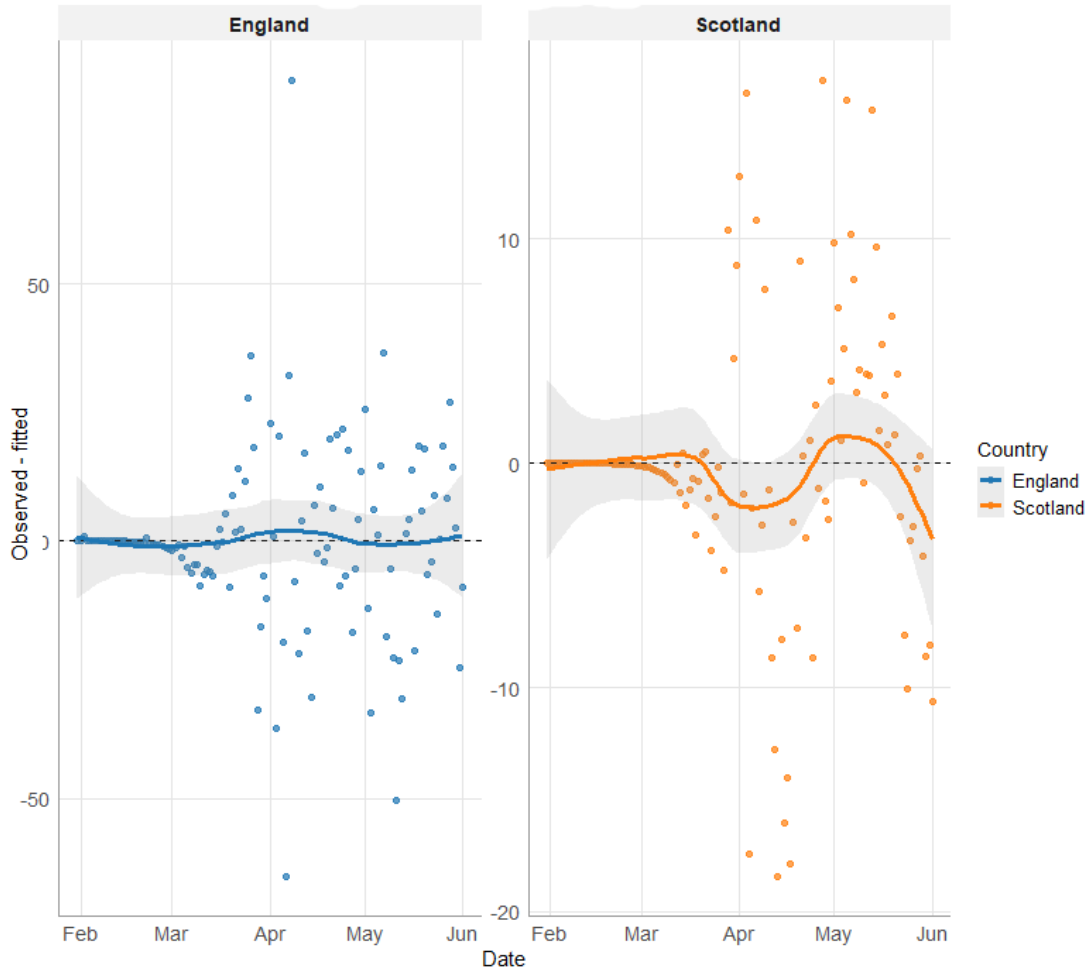


Figure 6.11: Residuals over time for the hierarchical logistic model refitted with infection-to-death delay $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$, shown for England and Scotland. Points are daily residuals, computed as observed minus fitted deaths, and smooth curves summarise the residual trend over time. Compared with the baseline logistic fit in Figure 6.5, the main systematic residual pattern is reduced in England, especially around the peak and the post-lockdown decline, while the change is smaller for Scotland.

the inferred contact floor and heterogeneity both move upward. At the country level, both England and Scotland move toward higher fitted transmission and weaker maximal contact reduction using the narrower-delay refit.

This sensitivity analysis shows that the remaining England discrepancy is not only a transmission-side issue about the choice of NPI shape. The observation layer, especially the spread of the infection-to-death delay, feeds back into the inferred transmission dynamics. Changing the delay distribution does not merely sharpen the fitted death curve; it also changes how the model allocates explanatory weight across $\mathcal{R}_0$, contact reduction, intervention timing, residual variation, and susceptibility heterogeneity. This is why the observation model is treated as a substantive part of the empirical inference, not simply as a technical conversion from infections to deaths.

Table 6.7: Key parameter changes when the hierarchical logistic model is refitted with the narrower infection-to-death delay $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$. Entries are posterior medians with central 95% credible intervals in brackets. The baseline logistic values are from Tables 6.2 and 6.3; the narrower-delay values are from the refit reported in Appendix G.

Parameter	Baseline logistic fit	Narrower-delay logistic refit
$\mathcal{R}_{0,\text{pop}}$	2.93 [2.22, 4.247]	3.03 [2.10, 4.288]
c_{pop}^*	0.264 [0.185, 0.403]	0.311 [0.211, 0.439]
ν	1.03 [0.479, 1.56]	1.84 [1.28, 2.30]
$\mathcal{R}_{0,\text{EN}}$	3.02 [2.96, 3.08]	3.20 [3.13, 3.27]
c_{EN}^*	0.246 [0.233, 0.268]	0.295 [0.266, 0.329]
$\mathcal{R}_{0,\text{SC}}$	2.70 [2.63, 2.78]	2.86 [2.78, 2.96]
c_{SC}^*	0.274 [0.250, 0.297]	0.324 [0.291, 0.356]

6.8 Comparison fits using uniform priors

Section 6.6.2 presented the main joint hierarchical Bayesian analysis, in which country-level parameters are regularised through partial pooling. To assess whether the main conclusions depend on that prior structure, we also consider a set of comparison fits with bounded uniform priors. These comparison fits use the same model, the same infection-to-death convolution, and the same likelihood and posterior checks as in Section 6.6.1. The difference being the uniform-prior fits do not introduce population-level hyper-parameters and do not apply partial pooling across countries. Instead, country-specific parameters are estimated directly on their own scales with broad bounded priors.

6.8.1 Model specification for uniform priors

For the uniform-prior comparison fits, we write

$$\theta^{(m)} = (\theta_{\text{EN}}^{(m)}, \theta_{\text{SC}}^{(m)}, \nu), \quad m \in \{\text{lin}, \text{log}, \text{str}\},$$

where ν is the susceptibility heterogeneity parameter shared across England and Scotland, and $\theta_j^{(m)}$ denotes the country-specific parameter vector for model m as explained in section 6.6.2. The parameter vectors for **piecewise linear**, **logistic** and **stringency-driven** are the same except that there is no partial pooling.

Appendix G reports all the plots of these comparison fits, including posterior predictive trajectories and inferred contact profiles for England and Scotland (Figures G.9–G.10), residual diagnostics Figure G.7.

Tables G.7 shows the uniform priors used in the fittings. We discuss the posterior summaries

below.

6.8.2 Posterior summaries of uniform prior fits

We summarise the posterior distributions in Tables 6.8, 6.9.

Table 6.8: Posterior summaries for the shared heterogeneity parameter (ν) for uniform priors. Entries are posterior medians with posterior standard deviation (SD) in parentheses and central 95% credible intervals in brackets.

Parameter	Linear	Logistic	Stringency
ν (centred)	0.359 (0.338) [0.0260, 1.26]	0.285 (0.252) [0.0224, 0.933]	1.59 (0.760) [0.109, 2.80]

Table 6.9: Country-level posterior summaries for fit with uniform priors for the three NPI specifications. Entries are posterior median with posterior standard deviation (SD) in parentheses and central 95% credible intervals in brackets. Parameters not present in a given NPI specification are shown as “—”.

Parameter	Posterior summaries (Q50 (SD) [Q2.5, Q97.5])		
	Piecewise linear	Logistic	Stringency
England			
$\mathcal{R}_{0,EN}$	3.36 (0.0281) [3.29, 3.40]	3.34 (0.0199) [3.29, 3.37]	3.29 (0.0708) [3.16, 3.43]
c_{EN}^*	0.203 (0.00627) [0.197, 0.221]	0.212 (0.00381) [0.208, 0.222]	0.228 (0.0329) [0.193, 0.311]
ϕ_{EN}	94.9 (22.3) [61.6, 148]	232 (76.0) [138, 422]	46.0 (10.7) [29.4, 71.7]
$I_{0,EN}$	10.5 (0.715) [10.0, 12.6]	10.4 (0.590) [10.0, 12.1]	17.4 (3.97) [11.0, 26.2]
$t_{0,EN}$	39.7 (0.661) [38.4, 41.0]	—	—
w_{EN} (days)	16.3 (0.661) [15.0, 17.6]	—	—
$t_{m,EN}$	—	48.2 (0.200) [47.8, 48.6]	—
k_{EN}	—	1.67 (0.264) [1.04, 1.99]	—
Scotland			
$\mathcal{R}_{0,SC}$	2.43 (0.0206) [2.38, 2.47]	2.42 (0.0290) [2.35, 2.46]	2.67 (0.0210) [2.62, 2.70]
c_{SC}^*	0.303 (0.00961) [0.287, 0.325]	0.295 (0.0103) [0.274, 0.315]	0.331 (0.0339) [0.292, 0.412]
ϕ_{SC}	62.0 (144) [24.8, 638]	48.9 (100) [21.2, 339]	31.3 (69.2) [14.1, 121]
$I_{0,SC}$	10.6 (0.824) [10.0, 13.0]	10.7 (1.04) [10.0, 13.8]	10.4 (0.561) [10.0, 12.1]
$t_{0,SC}$	53.0 (0.855) [50.8, 53.9]	—	—
w_{SC} (days)	1.00 (0.855) [0.0694, 3.18]	—	—
$t_{m,SC}$	—	53.9 (0.760) [52.5, 55.5]	—
k_{SC}	—	1.41 (0.423) [0.516, 1.97]	—

ϕ_j is the Negative-Binomial dispersion for country j . For the piecewise linear specification, $(t_{0,j}, w_j)$ describe onset and duration of contact reduction. For the logistic specification, $(t_{m,j}, k_j)$ describe midpoint and steepness. For the stringency-driven specification, $c(t)$ is driven by the observed OxCGRt series and has no additional timing/shape parameters beyond c_j^* .

Table 6.10 summarises fit using RMSE, MAE and bias.

6.8.3 Comparison with the hierarchical model analysis

The comparison between the hierarchical model fits and the uniform-prior fits tests sensitivity of priors to the joint inference.

Table 6.10: Goodness-of-fit metrics for the three NPI specifications for uniform priors. Metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Bias.

Country	Model	RMSE	MAE	Bias
England	Linear	42.9	26.7	0.326
England	Logistic	31.2	17.8	1.30
England	Stringency	52.7	35.6	-8.71
Scotland	Linear	5.87	3.71	-0.0540
Scotland	Logistic	6.08	3.83	-0.0780
Scotland	Stringency	6.03	3.95	0.315

The most visible is the shared heterogeneity parameter ν . In the hierarchical analysis, the posterior median for ν is 1.21 for the piecewise linear NPI, 1.03 for the logistic and 1.39 for the stringency-driven , with 95% credible intervals 0.647 to 1.79, 0.479 to 1.56, and 0.770 to 1.98, respectively; see Table 6.2. With uniform priors, the corresponding medians become 0.359, 0.285, and 1.59, with 95% credible intervals 0.0260 to 1.26, 0.0224 to 0.933, and 0.109 to 2.80; see Table 6.8. So for the linear and logistic specifications, the posterior moves much closer to the homogeneous limit when uniform priors are used. For the stringency specification, the opposite happens: the median moves upward from 1.39 to 1.59, while the interval becomes much wider. In fact, for the stringency fit with uniform priors, the posterior for ν still stretches from 0.109 to 2.80, which covers most of the prior support $\mathcal{U}(0.01, 3)$ in Table G.7. This suggests that, for this specification, the data provide only weak information about ν once the hierarchical regularisation is removed.

The contact-reduction parameter c^* is more stable in the hierarchical analysis than in the uniform-prior comparison. For England, the hierarchical estimates are 0.245, 0.246, and 0.238 for the linear, logistic, and stringency models, while the uniform-prior values are 0.203, 0.212, and 0.228. For Scotland, the hierarchical estimates are 0.283, 0.274, and 0.283, whereas the uniform-prior values are 0.303, 0.295, and 0.331; see Tables 6.3 and 6.9. So the hierarchical fits keep c^* relatively close across NPI forms, while the uniform-prior fits allow more movement, especially for Scotland and especially for the stringency NPI specification.

A similar shift appears in $\mathcal{R}_0$. For England, the hierarchical medians are 2.99, 3.02, and 3.10, whereas the uniform-prior medians are 3.36, 3.34, and 3.29 respectively for linear, logistic and stringency. For Scotland, the hierarchical medians are 2.70, 2.70, and 2.97, whereas the uniform-prior medians are 2.43, 2.42, and 2.67; Tables 6.3 and 6.9. So the England–Scotland gap becomes larger in the uniform-prior fits, not simply less precise. Thus there are changes in posterior location as well as posterior spread.

The fit summaries tell a similar story. For England, the logistic fit improves numerically when uniform priors are used: RMSE drops from 36.4 to 31.2, MAE from 21.7 to 17.8, and bias from 2.60 to 1.30. The linear fit also improves slightly, with RMSE changing from 48.1 to 42.9 and MAE from 30.4 to 26.7. But the stringency fit becomes clearly worse: RMSE rises from 47.4 to 52.7, MAE from 31.5 to 35.6, and bias moves from -2.74 to -8.71 . For Scotland, the differences are small in all three models: the linear RMSE changes from 5.91 to 5.87, the logistic RMSE from 6.00 to 6.08, and the stringency RMSE from 5.75 to 6.03; see Tables 6.1 and 6.10. So relaxing the prior does not produce a uniform gain in fit. What it mainly does is make the inference less stable across NPI forms.

These comparisons shows that the geometry of the priors in hierarchical analysis are better supported by the data and in Stan generally [101]. The uniform-prior fits are still useful as a sensitivity check and they connect naturally with the earlier applied fits in Chapter 3 which is plug-in-and-play.

6.8.4 Comparison with Gomes et al. [4]

Table 6.11: Comparison of main transmission parameters using a linear NPI specification. Entries are posterior medians with 95% credible intervals in brackets. For Gomes et al. [4], values are taken from Table 2 and their lockdown reduction parameter c_1 is reported here as c^* .

Parameter	Median [95% CrI]		
	Hierarchical	Uniform	Gomes et al. [4]
$\mathcal{R}_{0,EN}$	2.99 [2.92, 3.06]	3.36 [3.29, 3.40]	3.424 [3.285, 3.486]
$\mathcal{R}_{0,SC}$	2.70 [2.64, 2.77]	2.43 [2.38, 2.47]	3.365 [3.311, 3.460]
c_{EN}^*	0.245 [0.228, 0.274]	0.203 [0.197, 0.221]	0.2471 [0.2286, 0.2647]
c_{SC}^*	0.283 [0.262, 0.310]	0.303 [0.287, 0.325]	0.2471 [0.2286, 0.2647]
ν	1.21 [0.647, 1.79]	0.359 [0.0260, 1.26]	1.382 [0.6952, 1.762]

Note: The comparison is not fully like-for-like. In the present Chapter, deaths are estimated by convolving infections with a fixed distributed Log-normal infection-to-death delay, whereas Gomes et al. use a single estimated lag parameter in the infection–death link. Accordingly, differences in $\mathcal{R}_0$, c^* , and ν reflect both transmission/NPI assumptions and observation-model assumptions.

There is a methodological distinction between this chapter’s approach and Gomes et al. [4] : in the way observed deaths are linked to latent infections which makes direct comparison difficult. Parametrisations and initialisations are also different. Here, expected deaths are obtained via an explicit infection-to-death *delay convolution*: latent daily incidence is convolved with a discretised delay distribution $g(\ell)$ and scaled by the infection fatality ratio, with daily deaths modelled using overdispersed count noise as in Wood et al. [14]. By contrast, Gomes et al. [4] treat the infection-to-death timing through a *single* duration-offset parameter g , interpreted as the excess duration of fatal infections relative to non-fatal infections, and use this scalar delay directly when

constructing initial conditions from the first observed death count [4]. This difference matters because a distributed delay spreads deaths over multiple lags, which can alter the implied timing of peak incidence and therefore the trade-off between susceptibility heterogeneity and intervention effects in mortality-based inference. Section 6.7.5 makes this distinction more concrete.

Table 6.11 compares the numerical estimates for some common parameters in both. Relative to the uniform-prior fit, the hierarchical fit is clearly closer to Gomes et al. [4] in the inferred heterogeneity parameter. Using the piecewise linear NPI specification, the hierarchical model gives $\nu = 1.21 [0.647, 1.79]$, whereas the uniform-prior fit gives a much lower value, $\nu = 0.359 [0.0260, 1.26]$; Gomes et al. [4] report $\nu = 1.382 [0.6952, 1.762]$.

A similar pattern appears for England’s lockdown contact level. The hierarchical estimate, $c_{\text{EN}}^* = 0.245 [0.228, 0.274]$, is very close to the Gomes value, $0.2471 [0.2286, 0.2647]$, whereas the uniform-prior fit gives a lower value, $0.203 [0.197, 0.221]$.

6.9 Discussion and Conclusions

6.9.1 Summary results: hierarchical models

Across all three hierarchical NPI specifications, the fitted mortality trajectories reproduce the first-wave rise, peak, and decline reasonably well, with Scotland showing the cleanest overall fit (Figures 6.3–6.4). Among the three NPI forms, the logistic specification gives the best overall fit for England, with the lowest RMSE and MAE in Table 6.1 and also the best predictive performer in Table 6.4 in model selection (Section 6.7.4, Appendix G.3). For Scotland, the differences between NPI forms are much smaller, consistent with the smaller and noisier mortality series.

The model-based analysis also shows that inference is shaped not only by the transmission model, but also by the observation layer linking infections to deaths. Section 6.7.5 shows that the broader Wood’s [14] delay smooths and delays the implied death curve, whereas a narrower delay produces a sharper and better-aligned mortality peak. In the logistic refit with $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$, the England peak and post-lockdown decline are tracked more closely (Figures 6.10–6.11). However, this improvement changes the inferred parameters, with ν moving from 1.03 in the main logistic fit to 1.84, and c_{pop}^* moving from 0.264 to 0.311. Thus the remaining England discrepancy is not only about the transmission-side NPI shape; it also reflects sensitivity to the infection-to-death delay distribution.

6.9.2 Limitations and recommendations

The main limitation of this work is that it uses only two epidemics: England and Scotland. This should be distinguished from the simulation result in Chapter 3. In the controlled simulation setting, the two-epidemic design was tested under an idealised fully shared parameter vector $(\mathcal{R}_0, \nu, t_0, c^*)$, with epidemics differing mainly through their initial conditions. Under those assumptions, two well-contrasted epidemics were sufficient to recover most of the identifiability gain, narrowing the profile confidence interval for ν from a width of 0.666 to 0.039 (Table 3.1).

The empirical setting is more demanding. England and Scotland can be analysed under a common transmission and observation framework, but they are real epidemics with country-specific timing, reporting, intervention histories, and residual structure. The hierarchical modelling structure allows partial pooling between the two countries, providing a compromise between fitting them completely separately and forcing all parameters to be identical [101, 102]. However, with only two countries, the hierarchical layer is weakly informed about the distribution of country-level effects. In particular, between-country variance parameters and higher-level summaries cannot be learned with the same stability as they could be from a larger collection of regions or countries. This is a general limitation of multilevel models with very small numbers of higher-level units, where posterior inference can depend more strongly on the chosen hyperpriors and pooling assumptions [103, 104].

In the present analysis that dependence is not only a general concern; it is measurable directly. The comparison in Section 6.8.3 reveals how much of the inference the pooling structure is carrying. Using the hierarchical priors, the population-level posterior median for ν is 1.21 for the piecewise linear specification, 1.03 for the logistic, and 1.39 for the stringency-driven, and the three credible intervals share a common central 95% region of [0.770, 1.56] (Table 6.2). When the hierarchical regularisation is replaced by uniform priors, the same three specifications give medians of 0.359, 0.285, and 1.59 (Table 6.8), so the spread in ν across NPI forms widens. For the stringency specification the uniform-prior posterior for ν spans [0.109, 2.80], covering approximately 90% of the $U(0.01, 3)$ prior support, so the data carry almost no information about ν once the regularisation is removed. With two countries, the pooling structure is therefore doing a substantial part of the inferential work, and the stability seen in the hierarchical fits should be read as a property of the model as specified rather than as a property of the data alone.

The posterior summaries might therefore be interpreted as evidence that the joint framework can extract useful information from real mortality data, rather than as definitive population-level estimates of susceptibility heterogeneity. The estimates of ν , c^* , and $\mathcal{R}_0$ remain conditional on

the assumed observation model, NPI parameterisation, priors, and two-country pooling structure. Thus, the empirical analysis supports the feasibility of estimating heterogeneity alongside NPI effects, but it does not establish a stable distribution of country-level epidemic parameters.

A second limitation is that the first-wave mortality window contains only one major rise and decline in each country. This restricts the amount of information available about susceptible depletion, so some ν - c^* compensation remains unavoidable. The posterior dependence between ν and c^* is therefore not a nuisance feature to be ignored; it is part of the empirical result. The analysis shows that the joint model reduces some of the single-epidemic compensation, but it does not eliminate the underlying identifiability problem.

Another limitation concerns the observation layer. The data are deaths rather than infections, so latent infections must be mapped to observed mortality through an infection-to-death delay distribution and an infection fatality ratio. The sensitivity analysis in Section 6.7.5 shows that the remaining England misfit is partly affected by the assumed spread of the infection-to-death delay. In particular, using a narrower Lognormal delay distribution improves the fitted England mortality peak and post-lockdown decline, but also shifts the inferred values of ν and c^* . Thus, the delay distribution is not only a technical observation-model detail; it can feed back into the inferred transmission and heterogeneity parameters. Future sensitivity checks should therefore compare the Lognormal assumption with alternative delay specifications, such as Gamma, Weibull, or Erlang delay distributions, as well as empirical or non-parametric discrete delay kernels. A further extension would be to model the infection-to-death delay as the convolution of separate infection-to-onset and onset-to-death components, rather than as a single parametric delay distribution.

A fourth limitation is that the empirical fits retain the reduced Gamma-based heterogeneity structure. Chapter 5 showed that susceptibility-distribution misspecification can produce precise but biased inference. Future empirical work should therefore combine the present England–Scotland framework with the discretised susceptibility models developed earlier in the thesis, so that Gamma, Lognormal, and other positive susceptibility distributions can be compared directly.

Future work should extend the analysis to more countries or subnational regions, more epidemic waves, and additional data streams such as hospitalisations, serology, mobility, or contact data. Region-level covariates, such as age structure, population density, reporting differences, or mobility measures, could also be included so that some between-region variation is explained directly rather than absorbed entirely into random effects. A fully stochastic state-space

version of the empirical model would also be valuable, since the current analysis uses deterministic latent transmission dynamics with observation-level noise.

6.9.3 Conclusions

This chapter extended the thesis from controlled simulation studies to real-data joint inference for the first wave of COVID-19 mortality in England and Scotland. The main aim was to assess whether susceptibility heterogeneity can still be estimated meaningfully when transmission is modelled jointly with NPIs and when inference is based on delayed mortality data rather than directly observed infections. In this setting, the chapter connects directly to the identifiability results of Chapters 2 and 3: heterogeneity in susceptibility and NPIs effects can produce partially compensating effects on a single epidemic trajectory, but joint inference across related epidemics can still recover meaningful information when the model is constrained appropriately. In summary, the joint England–Scotland analysis supports three conclusions. First, among the three NPI specifications studied here, the logistic contact-reduction model gives the strongest predictive performance for England, while fit differences are smaller for Scotland. Second, the lockdown contact level c^* is comparatively stable across NPI forms in the hierarchical analysis, whereas the heterogeneity in susceptibility, ν , is moderately sensitive to the assumed NPI and observation-model structure: the credible intervals for ν overlap across all three specifications, but the posterior median shifts systematically with the assumed contact-transition shape. The assumed NPI form therefore does not determine whether heterogeneity is detected, but it does affect how much is inferred. Third, ν remains partially confounded with intervention strength, but it is still estimable to a useful degree using the hierarchical joint model. This third conclusion is conditional on the hierarchical structure: under uniform priors the linear and logistic fits move towards the homogeneous limit and the stringency fit becomes nearly uninformative about ν . In a nutshell, joint inference across England and Scotland makes susceptibility heterogeneity more empirically accessible than a single-epidemic analysis would be, and the posterior for ν remains informative using the fitted hierarchical models. However, the analysis also shows why conclusions from only two epidemics must remain cautious. The ν – c^* ridge is reduced but not removed; the results are sensitive to the assumed NPI form, to the observation model, and to the priors; and the mortality data do not observe infections directly. The empirical contribution of this chapter is therefore not to claim that heterogeneity has been conclusively identified, but to show how the joint-inference framework can be applied to real data while making the remaining uncertainty explicit.

Chapter 7

General Summary and Conclusion

7.1 Overview

This thesis studied inference for SEIR epidemic models with heterogeneous susceptibility and time-varying non-pharmaceutical interventions (NPIs). The main problem throughout was that both mechanisms can reduce transmission over time and can therefore produce similar epidemic curves. As a result, a model may fit observed data well without clearly revealing how much of the decline in transmission came from intrinsic population heterogeneity and how much came from changes in contact behaviour. The thesis addressed this problem in a sequence of linked steps. Chapter 1 established the mathematical and literature background. Chapter 2 examined what can be learned from a single epidemic time series. Chapter 3 asked whether joint analysis of more than one epidemic can improve identifiability. Chapter 4 developed discretisation tools for continuous susceptibility distributions so that the modelling framework is not restricted to the reduced Gamma formulation. Chapter 5 then used those tools to study robustness to distributional misspecification. Finally, Chapter 6 applied the framework to first-wave COVID-19 mortality data from England and Scotland.

In summary, the thesis shows that the main inferential difficulty is real, persistent, and practically important. At the same time, it also shows that the problem is not hopeless. Careful model design, joint analysis across related epidemics, and explicit robustness checks can recover useful information, while also making the remaining uncertainty more transparent.

7.2 Summary of the main findings

7.2.1 Single-epidemic inference has clear practical limits

Chapter 2 showed that, under the reduced heterogeneous-susceptibility SEIR model, the key parameter vector

$$(\mathcal{R}_0, \nu, t_0, c^*)$$

can be practically difficult to estimate from a single noisy epidemic when susceptibility heterogeneity and NPIs are both active. The problem we are interested in is the compensation between ν and c^* : stronger heterogeneity can be offset by weaker intervention, and vice versa, while producing very similar fitted incidence curves. A related trade-off also appears between $\mathcal{R}_0$ and t_0 .

An important message from Chapter 2 is that this is not just a numerical accident. It arises because, over a single epidemic time series, the observed trajectory mainly constrains a composite effect on transmission rather than uniquely separating the underlying mechanisms. This means that apparently good fits do not by themselves imply reliable parameter recovery.

The appendices to Chapter 2 strengthen this point. For instance, Appendix B shows that the ν - c^* dependence persists across estimation methods and across different NPI functional forms. The tau-leaping appendix then shows that the same compensation remains under fully stochastic epidemic simulation. In large populations, the stochastic results closely agree with the deterministic analysis, while in smaller populations the main effect of process noise is to widen uncertainty rather than remove the ridge.

7.2.2 Joint-epidemic inference can substantially improve identifiability

Chapter 3 showed that When two epidemics series are analysed jointly under shared parameters, the second epidemic acts like an additional constraint. Parameter combinations that fit one epidemic must also fit the other, and this sharply reduces the ν - c^* compensation seen in the single-epidemic case.

This improvement was visible in several ways: narrower profile-likelihood intervals, better Hessian conditioning, weaker parameter correlations, and more compact likelihood surfaces. The gain was especially strong for ν and c^* , which were the most strongly confounded in Chapter 2. For ν , the profile confidence interval narrowed from a width of 0.666 to 0.039.

However, Chapter 3 also showed that the improvement has limits, and that those limits are set by the sharing assumptions rather than by the number of epidemics alone. The gain was

obtained under an idealised design in which the whole vector $(\mathcal{R}_0, \nu, t_0, c^*)$ was shared exactly and the epidemics differed only in their initial conditions. When more parameters are allowed to vary by epidemic, the compensation begins to re-emerge. Likewise, adding a third epidemic gave only modest extra benefit relative to the move from one epidemic to two. Therefore, the main lesson is not that joint-epidemic analysis solves every identifiability problem, but that it can help strongly when the sharing assumptions are scientifically reasonable. This condition is exactly what real epidemics strain, and Section 7.4 returns to it.

In particular, varying $\mathcal{R}_0$ within the plausible range 2.5 to 3.5 (Appendix A.4) does not alter the qualitative behaviour of the model, nor the comparative conclusions regarding discretisation schemes and NPI parametrisations. This indicates that the results are robust to moderate uncertainty in the baseline transmission intensity.

7.2.3 Discretisation is necessary for moving beyond the reduced Gamma model

A further contribution of the thesis was computational. The reduced Gamma model is attractive because it gives a low-dimensional closed form, but this convenience does not directly extend to other continuous susceptibility distributions. Chapter 4 addressed this by developing a discretisation framework for continuous heterogeneity.

The chapter introduced a distribution-agnostic moment-calibration step and then specialised it to the Gamma and Lognormal cases. The main goal was to preserve the first two moments of the distributions while producing a positive, ordered, and numerically stable discrete approximation that can be used inside the epidemic model. Validation was carried out at two levels: first at the level of the probability distribution itself, and then at the level of epidemic trajectories.

This part of the thesis is important because it provides the numerical foundation needed for the misspecification analysis in Chapter 5 and for future work with other susceptibility families. In that sense, Chapter 4 is not separate from the identifiability story. It is the computational bridge that allows the thesis to ask whether the conclusions are robust to the assumed family of heterogeneity.

7.2.4 Distributional shape matters once heterogeneity is moderate to high

Chapter 5 showed that matching only the first two moments of the susceptibility distribution is not enough to guarantee the same epidemic trajectory. When heterogeneity is moderate to high, higher-order shape features become important. In the comparisons studied here, the Lognormal

distribution produced later and larger peaks and larger final sizes than the Gamma distribution, even when both had the same mean and coefficient of variation.

This has two important inferential consequences. First, single-epidemic goodness-of-fit can hide misspecification. A misspecified model can often absorb the structural error through a compensating shift in ν , while leaving the fitted trajectory visually acceptable over the training window. Second, the forecast can still be wrong in a systematic way. In this thesis, fitting a lighter-tailed family to heavier-tailed baseline tended to give an optimistic forecast, while fitting a heavier-tailed family to lighter-tailed baseline tended to give a pessimistic one. However, the forecast cost of choosing the wrong heterogeneous family is smaller than the cost of ignoring heterogeneity altogether: the homogeneous reference overshoots peak incidence by hundreds of percent where the misspecified heterogeneous fit stays within a few percent to a few tens of percent.

Chapter 5 also showed that joint-epidemic inference does not remove structural misspecification. Instead, it can expose it more clearly. When the susceptibility family is correctly specified, the sharper identifiability from two-epidemic fitting is beneficial. When the family is wrong, the same sharpening can force the model toward a precise but biased explanation. This is an important caution: improved identifiability is only helpful when the model class is adequate. From a forecasting point of view, however, the asymmetry was milder under Lognormal baseline: the misspecified Gamma model still gave reasonably good two-epidemic forecasts in the low-to-moderate heterogeneity range, with only moderate underestimation of the second-wave peak at $\nu = 1.414$, whereas the reverse Lognormal misspecification under Gamma baseline produced a much stronger forecast overshoot.

7.2.5 The empirical England–Scotland analysis supports partial but not complete separation

Chapter 6 brought the thesis to real data by fitting the heterogeneous framework to first-wave COVID-19 mortality in England and Scotland. This moved the problem from idealised simulation into a more realistic setting with observation delays, IFR assumptions, initial-condition uncertainty, and alternative NPI parametrisations. Bayesian inference techniques were used throughout the chapter.

The main empirical message is balanced. On one hand, the ridge does not disappear in real data. Posterior dependence between heterogeneity and intervention strength remains visible, especially in England. On the other hand, joint and hierarchical fitting still recover useful

information. Across the NPI specifications studied, the intervention floor c^* was relatively stable, the logistic specification gave the best fit among the tested NPI forms, and ν remained estimable to a meaningful though still imperfect degree.

The empirical chapter therefore supports the general thesis conclusions. It suggests that heterogeneity cannot be read off from a single first-wave mortality curve without caution. But it also shows that, with joint analysis of related epidemics and with appropriate regularisation, the posterior for ν can still be informative enough to support scientific comparison across model specifications. The qualification carried by the word regularisation is substantial, and Chapter 6 quantifies it. Under the hierarchical priors the posterior median for ν lies between 1.03 and 1.39 across the three NPI specifications, and all three credible intervals exclude zero. When the hierarchical structure is replaced by uniform priors, the linear and logistic specifications collapse towards the homogeneous limit, with medians of 0.359 and 0.285, while the stringency-driven posterior stretches across $[0.109, 2.80]$, covering roughly 90% of its prior support and therefore carrying almost no information about ν . With two countries, the pooling structure is doing a substantial part of the inference rather than merely stabilising it, and this is the direct reason why the empirical conclusions of Chapter 6 are stated conditionally.

7.3 Main contributions of the thesis

The contributions of this thesis can be summarised in five parts. First, it gave a clear and detailed practical identifiability analysis for the closed form Gamma distributed susceptibility SEIR model under NPIs, with explicit focus on the compensation mechanisms among $(\mathcal{R}_0, \nu, t_0, c^*)$. Second, it showed that the strongest ridge in single-epidemic inference is the ν – c^* trade-off, and that this effect is robust across estimation methods, NPI functional forms, and even stochastic tau-leaping simulation. The joint inference still has a trade-off in another pair of correlations, on $\mathcal{R}_0$ – t_0 , which persists in both single-epidemic and two-epidemic fittings. Third, it demonstrated that joint-epidemic inference can materially improve identifiability when suitable sharing assumptions hold, while also clarifying the limits of this gain under higher-dimensional extensions and additional epidemics. Fourth, it developed a practical discretisation framework for continuous susceptibility distributions, making it possible to study heterogeneous epidemic models beyond the reduced Gamma form in derived in Section 1.6.3. Fifth, it connected theory, simulation, computation, and data by applying the framework to real COVID-19 mortality data from England and Scotland, thereby showing both what can be learned empirically using our methods and what remains sensitive to modelling assumptions. This fifth contribution is a demonstration of feasibility on

two epidemics, not an estimate of a population-level distribution of susceptibility heterogeneity, and it should be read as such.

7.4 Limitations of the present work

There are several limitations that should be kept in mind, and one of them is more consequential than the rest.

The main limitation of this work is that its empirical demonstration rests on two epidemics only, England and Scotland, and this is what generates the uncertainty attached to the conclusions drawn from it. The reason is not simply that two is a small number. It is that the assumption which made two epidemics sufficient in simulation does not hold in the data. Chapter 3 obtained its identifiability gain under a design in which the full parameter vector $(\mathcal{R}_0, \nu, t_0, c^*)$ was shared exactly across epidemics that differed only in their initial conditions. England and Scotland do not satisfy that condition. They are real epidemics with country-specific timing, reporting, intervention histories, and residual structure, so the empirical model in Chapter 6 requires a hierarchical layer that allows partial pooling instead of exact sharing.

With only two countries, that hierarchical layer is weakly informed about the distribution of country-level effects. Between-country variance parameters and higher-level summaries cannot be learned with the stability that a larger collection of regions would allow, so the hyperpriors and the pooling assumptions carry a substantial part of the inference rather than merely regularising it. Chapter 6 measures the size of that contribution rather than asserting it: removing the hierarchical regularisation widens the spread of ν across the three NPI specifications and leaves the stringency-driven fit nearly uninformative about ν . Every empirical estimate reported in Chapter 6 is therefore conditional on the two-country pooling structure, and none of them establishes a general distribution of country-level epidemic parameters. This limitation cannot be repaired by refining the model. It requires more epidemics.

The remaining limitations are narrower. The simulation studies focused on a reduced SEIR structure and on a restricted set of parametrisations for intervention effects. Although this was necessary for clarity and tractability, real epidemics may involve additional sources of heterogeneity, such as age, space, contact-network structure, and reporting variation, which were not modelled explicitly here. Moreover, the parameter choices are primarily motivated by the COVID-19 pandemic. While this provides a well-documented and relevant case study, other infectious diseases may exhibit different transmission characteristics. For example, seasonal influenza typically has a lower basic reproduction number but shorter generation times, while

emerging pathogens may exhibit higher growth rates. Future work could extend the present framework to other diseases with different epidemiological profiles, as well as to longer epidemic time horizons (for example, multi-year dynamics), where additional factors such as waning immunity and demographic turnover become important.

The misspecification analysis in Chapter 5 focused mainly on the Gamma and Lognormal comparison. This was a reasonable starting point, especially because the Lognormal gives a direct heavy-tail alternative to Gamma, but it does not cover other biologically plausible susceptibility families.

The empirical analysis also depends on its observation layer. The information content of a first-wave death series is limited, since it contains only one major rise and decline in each country, so some degree of confounding between ν and c^* is unavoidable regardless of the number of epidemics. The analysis further depends on assumptions about the infection-to-death delay and the IFR, which can influence the inferred mortality curve even when the underlying transmission model is unchanged, and it uses a single delay family only, the Lognormal. Testing the sensitivity of the results to alternative delay distributions would be worthwhile.

Finally, although stochastic dynamics were examined in the tau-leaping in section 2.7, the main empirical inference remained based on a deterministic latent epidemic model linked to noisy observations. A fully stochastic state-space treatment of the empirical problem was beyond the present scope.

7.5 Directions for future work

Several extensions would be valuable.

The most important next step follows directly from the main limitation above. Extending the empirical framework to more regions, more waves, or multiple data streams such as deaths, hospitalisations, serology, and mobility data would give the hierarchical layer enough higher-level units to be informed by the data rather than by its hyperpriors. Such designs would also provide stronger and more diverse information for separating susceptible depletion effects from intervention effects, and would allow a direct test of whether the England and Scotland findings generalise. Region-level covariates such as age structure, population density, reporting differences, or mobility could be included so that some between-region variation is explained directly rather than absorbed into random effects.

A second direction is to broaden the family of susceptibility distributions studied. The discretisation framework developed in Chapter 4 makes this feasible. This would allow systematic

comparison of additional heavy-tailed or skewed families and a clearer understanding of how much inference depends on the assumed tail behaviour.

A third direction is to combine the present identifiability framework with richer epidemic structure, for example age structure, spatial coupling, contact heterogeneity, reinfection, or vaccination. These extensions would make the model more realistic, but they would also raise new identifiability questions. The lesson from this thesis is that added realism should always be evaluated together with the information needed to support it.

A fourth direction is to develop stochastic empirical inference more fully. The tau-leaping section in Chapter 2 suggests that process noise mainly widens uncertainty rather than overturning the basic conclusions, but a full stochastic observation and transmission framework would allow that claim to be tested directly on data.

Lastly, the forecast results under misspecification suggest that identifiability and robustness should be studied together rather than separately. A model can be sharply identified and still be wrong. Future work should therefore place more emphasis on joint assessment of fit, forecast performance, and structural adequacy.

7.6 Final conclusion

The main conclusion of this thesis is simple. In epidemic models with heterogeneous susceptibility and NPIs, good fit to a single epidemic is not enough to guarantee trustworthy inference. The data may only identify a compensating combination of mechanisms rather than the mechanisms themselves. This is why heterogeneity and intervention strength are so difficult to separate from single epidemic data series alone.

At the same time, the thesis also shows a constructive way forward. Joint-epidemic inference can improve identifiability when common structure across epidemics is justified. Discretisation methods make it possible to move beyond the reduced Gamma model and test robustness more seriously. Misspecification analysis reveals that sharper inference is only valuable when the model family is adequate. The England and Scotland application shows that, even in real data, useful information can still be recovered if the remaining uncertainty is made explicit.

That constructive claim should be read at the scale on which it was demonstrated. The framework was applied to two epidemics, and two epidemics are enough to show that the approach works while remaining too few to inform the hierarchical structure the approach relies on. This is the principal reason the empirical conclusions of Chapter 6 are stated conditionally, and it is a limitation of the evidence rather than of the method.

Appendix A

Mathematical and Statistical Details

A.1 Reference notation for the heterogeneous-susceptibility SEIR model

This appendix collects notation and auxiliary calculations used across the thesis. The detailed Gamma-reduction and shape-preservation arguments have been moved to Chapter 1, especially Sections 1.6.3 and 1.6.4. They are therefore not repeated here.

A.1.1 Continuous distribution model

Let $x > 0$ denote individual susceptibility. Write $S(x, t)$, $E(x, t)$, $I(x, t)$, and $R(x, t)$ for the densities or counts of hosts with susceptibility x in each epidemiological compartment at time t .

The aggregate compartment sizes are

$$S(t) = \int_0^\infty S(x, t) dx, \quad E(t) = \int_0^\infty E(x, t) dx, \quad I(t) = \int_0^\infty I(x, t) dx, \quad R(t) = \int_0^\infty R(x, t) dx,$$

and the total population size is

$$N = S(t) + E(t) + I(t) + R(t) = \int_0^\infty \{S(x, t) + E(x, t) + I(x, t) + R(x, t)\} dx.$$

Thus $S(t)$ is the integral of $S(x, t)$ itself. The density $q(x)$ denotes the initial susceptibility distribution and is normalised so that

$$\int_0^\infty q(x) dx = 1, \quad \int_0^\infty xq(x) dx = 1.$$

The second normalisation fixes the scale of x , so that the initial population-average susceptibility is one. If the epidemic seed is small, one may take $S(x, 0) \approx Nq(x)$; more generally $S(x, 0) = S_0q(x)$.

The force of infection is

$$\lambda(t) = \frac{\beta(t)}{N}(\rho E(t) + I(t)),$$

where $\rho \in [0, 1]$ is the relative infectiousness of exposed individuals and $\beta(t)$ may vary in time through a contact multiplier $c(t)$. The susceptibility-indexed SEIR system is

$$\begin{aligned} \frac{\partial S(x, t)}{\partial t} &= -x\lambda(t)S(x, t), \\ \frac{\partial E(x, t)}{\partial t} &= x\lambda(t)S(x, t) - \delta E(x, t), \\ \frac{\partial I(x, t)}{\partial t} &= \delta E(x, t) - \gamma I(x, t), \\ \frac{\partial R(x, t)}{\partial t} &= \gamma I(x, t). \end{aligned} \tag{A.1}$$

When NPIs act, transmission is scaled as

$$\beta(t) = \frac{\mathcal{R}_0 c(t)}{\rho/\delta + 1/\gamma}, \quad \mathcal{R}(t) = c(t)\mathcal{R}_0,$$

so $c(t) \in (0, 1]$ changes the level of transmission without changing the latent and infectious time scales δ^{-1} and γ^{-1} .

A.1.2 Reduced Gamma reference form

For a Gamma susceptibility distribution with mean one and coefficient of variation ν , the continuous system (A.1) reduces to the aggregate Gamma model derived in Chapter 1. In the count-scale notation used in the simulation chapters, the reference form is

$$\begin{aligned} \frac{dS}{dt} &= -\beta(t)(\rho E + I) \left(\frac{S}{N} \right)^{1+\nu^2}, \\ \frac{dE}{dt} &= \beta(t)(\rho E + I) \left(\frac{S}{N} \right)^{1+\nu^2} - \delta E, \\ \frac{dI}{dt} &= \delta E - \gamma I, \\ \frac{dR}{dt} &= \gamma I. \end{aligned} \tag{A.2}$$

The homogeneous SEIR model is recovered by setting $\nu = 0$. The parameter $\mathcal{R}_0$ is unchanged by the variance of susceptibility because the initial susceptibility distribution is normalised to have mean one; heterogeneity affects later dynamics through selective depletion and the factor $(S/N)^{1+\nu^2}$.

A.1.3 Initial conditions: the E/I ratio in the early phase

In the early epidemic ($S/N \approx 1$, with no intervention so $c(t) \equiv 1$), the (E, I)-subsystem of (A.2) linearises to

$$\frac{d}{dt} \begin{pmatrix} E \\ I \end{pmatrix} = \underbrace{\begin{pmatrix} \rho\beta - \delta & \beta \\ \delta & -\gamma \end{pmatrix}}_{=: A} \begin{pmatrix} E \\ I \end{pmatrix}, \quad \beta = \frac{\mathcal{R}_0}{\rho/\delta + 1/\gamma}. \quad (\text{A.3})$$

Eigenvalues. The characteristic polynomial of A is

$$\lambda^2 + \lambda(\gamma + \delta - \rho\beta) - [(\rho\beta - \delta)\gamma + \beta\delta] = 0,$$

giving

$$\lambda_{\pm} = \frac{(\rho\beta - \delta - \gamma) \pm \sqrt{(\rho\beta - \delta + \gamma)^2 + 4\delta\beta}}{2}. \quad (\text{A.4})$$

Eigenvector and the E/I ratio. For the dominant eigenvalue λ_+ , the eigenvector equation $(A - \lambda_+ I)v = 0$ yields from the second row $\delta E + (-\gamma - \lambda_+)I = 0$, so

$$\frac{E}{I} = \frac{\gamma + \lambda_+}{\delta} = \frac{(\rho\beta - \delta + \gamma) + \sqrt{(\rho\beta - \delta + \gamma)^2 + 4\delta\beta}}{2\delta}. \quad (\text{A.5})$$

This expression can be evaluated from $(\mathcal{R}_0, \delta, \gamma, \rho)$ alone using $\beta = \mathcal{R}_0/(\rho/\delta + 1/\gamma)$. The practical initialisation rule is therefore

$$\boxed{E_0 = \frac{\gamma + \lambda_+}{\delta} I_0, \quad \lambda_+ \text{ from (A.4).}}$$

Numeric example (baseline used throughout). With $\delta = 1/5.5$, $\gamma = 1/4$, $\rho = 0.5$, $\mathcal{R}_0 = 3$:

$$\beta = \frac{3}{0.5 \times 5.5 + 4} = \frac{3}{6.75} \approx 0.4444 \text{ day}^{-1}, \quad \lambda_+ \approx 0.214 \text{ day}^{-1}, \quad \frac{E}{I} \approx 2.554.$$

For simplicity and reproducibility we adopt $E_0 \approx 2.5 I_0$ throughout. This ratio is independent of heterogeneity ν because $S/N \approx 1$ at the start.

A.1.4 Computing $\mathcal{R}_0$ for system (A.1)

We use a discrete approximation of the susceptibility distribution. Partition the population into n susceptibility classes, where class i has susceptibility $x_i > 0$ and population proportion $p_i > 0$ with $\sum_i p_i = 1$. At total population N , class i contains Np_i individuals. Let $S_i(t)$, $E_i(t)$, $I_i(t)$ be the susceptible, exposed, and infectious counts in class i .

The force of infection for class i is

$$\lambda_i(t) = \beta x_i \frac{\rho \sum_{j=1}^n E_j(t) + \sum_{j=1}^n I_j(t)}{N},$$

where δ is the $E \rightarrow I$ rate, γ the recovery rate, and ρ the relative infectiousness of exposed individuals. Heterogeneity acts only on *susceptibility to infection*, not on infectiousness.

Class-specific system. For $i = 1, \dots, n$:

$$\dot{S}_i = -\beta x_i \frac{\rho \sum_j E_j + \sum_j I_j}{N} S_i, \quad (\text{A.6})$$

$$\dot{E}_i = \beta x_i \frac{\rho \sum_j E_j + \sum_j I_j}{N} S_i - \delta E_i, \quad (\text{A.7})$$

$$\dot{I}_i = \delta E_i - \gamma I_i. \quad (\text{A.8})$$

Disease-free equilibrium (DFE). At the DFE, $E_i^* = I_i^* = 0$ and $S_i^* = Np_i$ for all i : no class has been depleted by infection yet. The mean susceptibility at the DFE is

$$\bar{x} := \sum_{i=1}^n p_i x_i.$$

Selective depletion across classes appears only after transmission begins, so $\bar{x}$ is the relevant quantity for $\mathcal{R}_0$.

Linearisation at the DFE. Replace S_i by Np_i in the new-infection term of (A.7), and define totals $E = \sum_i E_i$ and $I = \sum_i I_i$. Summing (A.7)-(A.8) over i :

$$\dot{E} = \beta \bar{x} \rho E + \beta \bar{x} I - \delta E, \quad (\text{A.9})$$

$$\dot{I} = \delta E - \gamma I. \quad (\text{A.10})$$

Next-generation decomposition. Write $\dot{\mathbf{z}} = (F - V)\mathbf{z}$ with $\mathbf{z} = (E, I)^\top$. The new-infection Jacobian and transition Jacobian at the DFE are

$$F = \begin{pmatrix} \beta \bar{x} \rho & \beta \bar{x} \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \delta & 0 \\ -\delta & \gamma \end{pmatrix}.$$

Inverting V (with $\det V = \delta\gamma$):

$$V^{-1} = \frac{1}{\delta\gamma} \begin{pmatrix} \gamma & 0 \\ \delta & \delta \end{pmatrix}.$$

The next-generation matrix is

$$K = FV^{-1} = \frac{1}{\delta\gamma} \begin{pmatrix} \beta\bar{x}(\rho\gamma + \delta) & \beta\bar{x}\delta \\ 0 & 0 \end{pmatrix}.$$

K is upper triangular, so its eigenvalues are the diagonal entries: $\lambda_1 = \beta\bar{x}(\rho/\delta + 1/\gamma)$ and $\lambda_2 = 0$. Therefore, the basic reproduction number is the spectral radius, denoted here by $\varrho(K)$ to avoid confusion with the exposed-class relative infectiousness parameter ρ :

$$\boxed{\mathcal{R}_0 = \varrho(K) = \beta\bar{x} \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right)}. \quad (\text{A.11})$$

Interpretation. Under this model, heterogeneity enters only through susceptibility to *acquiring* infection. $\mathcal{R}_0$ therefore depends on susceptibility only through its *mean value at the DFE*, $\bar{x}$. If the distribution is normalised to unit mean ($\bar{x} = 1$), then

$$\mathcal{R}_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right),$$

which is identical to the homogeneous expression in equation (1.12). The variance (or coefficient of variation ν) of susceptibility is invisible to the initial growth rate; it manifests only through the time-varying effective reproduction number $\mathcal{R}_{\text{eff}}(t)$ as selective depletion progressively lowers the mean susceptibility of the remaining susceptible pool.

Consistency with the reduced system. Setting $S/N = 1$ and $c(t) \equiv 1$ in (A.2) and applying the same next-generation argument to the (E, I) subsystem gives

$$F = \begin{pmatrix} \rho\beta & \beta \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \delta & 0 \\ -\delta & \gamma \end{pmatrix}, \quad \mathcal{R}_0 = \varrho(FV^{-1}) = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right),$$

confirming that (A.11) holds equally for the reduced model. Under interventions, $\mathcal{R}(t) = c(t)\mathcal{R}_0$.

A.2 Three NPI functional forms and window matching

Chapter 1, section 1.7 introduced parametric specification for non-pharmaceutical interventions (NPIs) for this thesis. Chapters 2, 3, 5 use the piecewise-linear profile as the baseline with alternative NPI profiles in robustness and identifiability checks in Appendix B for Chapter 2. Chapter 6 uses logistic and stringency-driven profiles in the empirical analysis in addition to the

baseline piecewise-linear for the thesis. This section defines the three smooth functional forms mathematically, derives the parameter relationships that make them cover approximately the same transition window $[t_0, t_1]$, and shows that the three curves behave qualitatively similarly once their parameters are calibrated to share the same onset, target level, and window width. The same parameter values are used in the robustness analysis of Appendix B, so this section serves as the single mathematical reference for both.

A.2.1 The three functional forms

Let $c(t) \in (0, 1]$ denote the contact multiplier, where $c = 1$ is the baseline and $c^* \in (0, 1)$ is the sustained post-intervention level. All three forms share:

- t_0 : time of first behavioural change (onset of voluntary contact reduction),
- c^* : long-run contact level (NPI strength parameter),
- $w = t_1 - t_0$: transition window width.

Form 1: Piecewise-linear. [4, 62]

$$c_{\text{lin}}(t) = \begin{cases} 1, & t \leq t_0, \\ 1 - (1 - c^*) \frac{t - t_0}{t_1 - t_0}, & t_0 < t \leq t_1, \\ c^*, & t > t_1. \end{cases} \quad (\text{A.12})$$

Transmission declines linearly from onset t_0 and reaches c^* exactly at t_1 . This is the baseline specification used throughout the thesis.

Form 2: Exponential decay. [62]

$$c_{\text{exp}}(t) = \begin{cases} 1, & t \leq t_0, \\ c^* + (1 - c^*) e^{-\kappa(t-t_0)}, & t > t_0, \end{cases} \quad (\text{A.13})$$

where $\kappa > 0$ is the decay rate. The multiplier declines exponentially from 1 and approaches c^* asymptotically; the transition window is characterised by how long it takes to complete a given fraction of the total drop.

Form 3: Logistic (sigmoid) transition. [62]

$$c_{\log}(t) = c^* + \frac{1 - c^*}{1 + \exp\{k(t - t_m)\}}, \quad (\text{A.14})$$

where $k > 0$ is the steepness and t_m is the inflection midpoint. At $t = t_m$, $c_{\log}(t_m) = (1 + c^*)/2$, halfway between baseline and floor. The adoption fraction $p(t) = \sigma(k(t - t_m))$ follows a standard diffusion-of-innovation curve.

A.2.2 Matching the transition window

We want each form to complete approximately the same fraction of its total drop within $[t_0, t_1]$. The criterion is that each curve reaches the same completion fraction p_H by t_1 and, where applicable, has completed fraction p_L by t_0 . We use the symmetric 5-95% convention: $p_L = 0.05$, $p_H = 0.95$.

Linear. The linear form reaches exactly c^* at t_1 by definition, so it achieves 100% completion at t_1 . The 50% completion time is $t_{50} = (t_0 + t_1)/2$.

Exponential (90th-percentile matching). Requiring $c_{\exp}(t_1) = c^* + 0.10(1 - c^*)$ (90% completion by t_1):

$$e^{-\kappa w} = 0.10 \quad \implies \quad \boxed{\kappa = \frac{\ln 10}{w}}.$$

For general target fraction $1 - \varepsilon$ at t_1 : $\kappa = -\ln(\varepsilon)/w$.

Logistic (5-95% quantile matching). Imposing $p(t_0) = p_L = 0.05$ and $p(t_1) = p_H = 0.95$:

$$k w = \text{logit}(p_H) - \text{logit}(p_L) \quad \implies \quad \boxed{k = \frac{\Delta_{\text{span}}}{w}, \quad \Delta_{\text{span}} = \text{logit}(p_H) - \text{logit}(p_L)}.$$

For symmetric $p_L = 1 - p_H$, the midpoint is $\boxed{t_m = (t_0 + t_1)/2}$ and $\Delta_{\text{span}} \approx 5.887$ for $p_L = 0.05$.

Summary. Table A.1 collects the parameter formulas for all three forms for general window width w , together with the concrete values for the window $w = 5$ days used in Appendix B and throughout the simulation chapters of this thesis ($t_0 = 15$, $t_1 = 20$, $c^* = 0.3$).

Table A.1: Parameters for three NPI functional forms matched to the same onset t_0 , window width $w = t_1 - t_0$, and target level c^* . The right column gives numerical values for the simulation setting $t_0 = 15$, $t_1 = 20$ ($w = 5$ days), $c^* = 0.3$, which is used in Appendix B and throughout the thesis simulation studies.

Form	Parameter	Window-matching formula	Value ($w = 5$)
Linear	$t_1 = t_0 + w$	Reaches c^* exactly at t_1 ; 50% at $(t_0 + t_1)/2$	$t_1 = 20$ days
Exponential	κ	$\kappa = \ln(10)/w$ (90% completion at t_1)	$\kappa \approx 0.143 \text{ day}^{-1}$
Logistic	k, t_m	$k = 5.887/w$; $t_m = (t_0 + t_1)/2$ (5-95% symmetric)	$k \approx 1.178 \text{ day}^{-1}$, $t_m = 17.5$ days

A.2.3 Numerical illustration

Figure A.1 illustrates all three forms with $t_0 = 15$ days, $t_1 = 20$ days (window width $w = 5$ days), and $c^* = 0.3$ (70% reduction). From Table A.1, the matched parameters are:

$$\text{Exponential: } \kappa = \ln(10)/5 \approx 0.143 \text{ day}^{-1},$$

$$\text{Logistic: } k = 5.887/5 \approx 1.178 \text{ day}^{-1}, \quad t_m = 17.5 \text{ days}.$$

All three curves share the same onset t_0 , the same long-run level c^* , and complete most of their transition within $[15, 20]$. The exponential declines fastest at onset because its rate of change is largest at t_0 , not at the midpoint. The logistic has its maximum rate of change at $t_m = 17.5$ days. The piecewise-linear form is the only one that reaches c^* exactly at t_1 ; the other two approach it asymptotically.

A.3 Basics of Bayesian Hierarchical Models

Bayesian hierarchical models extend standard Bayesian inference by introducing multiple layers of uncertainty. They are useful when observations are structured by groups, locations, or time periods, because they allow within-group variation and between-group variation to be represented in a single probability model [105].

Let y denote the observed data, $\theta = (\theta_1, \dots, \theta_J)$ the collection of group-specific parameters, and η the hyperparameters that describe the population-level distribution of the group parameters. Bayes' rule then gives the hierarchical posterior density, up to proportionality, as

$$p(\theta, \eta \mid y) \propto p(y \mid \theta) p(\theta \mid \eta) p(\eta). \quad (\text{A.15})$$

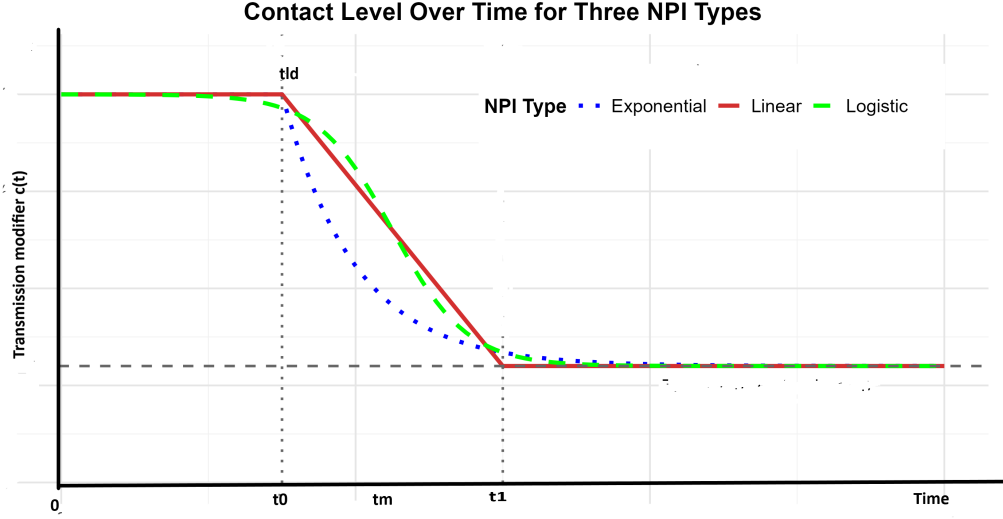


Figure A.1: Three NPI functional forms $c(t)$ sharing onset $t_0 = 15$ days, window $t_1 = 20$ days ($w = 5$ days), and target level $c^* = 0.3$ (70% contact reduction). **Linear** (red solid): piecewise-linear, equation (A.12), reaches c^* exactly at $t_1 = 20$. **Exponential** (blue dashed): decay with $\kappa = \ln(10)/5 \approx 0.143 \text{ day}^{-1}$, equation (A.13). **Logistic** (green dash-dot): sigmoid with $k \approx 1.178 \text{ day}^{-1}$ and midpoint $t_m = 17.5$ days, equation (A.14). The three profiles are synchronised using the window-matching formulas in Table A.1. The same parameter values are used in the robustness analysis of Appendix B.

The three factors in (A.15) correspond to the data model, the group-level model, and the hyperprior model, respectively.

Level 1 (data model / likelihood).

Let y_i denote observation i , and let $j(i)$ indicate the group to which observation i belongs. Conditional on the corresponding group-level parameter $\theta_{j(i)}$, the observation has probability mass or density function

$$p(y_i | \theta_{j(i)}). \quad (\text{A.16})$$

Thus the observations are conditionally independent given the relevant group-specific parameters.

Level 2 (group-level model).

Instead of estimating each group independently, the group-specific parameters θ_j are treated as unknown quantities generated from a shared population model governed by hyperparameters η . Equivalently, conditional on η , θ_j has probability density

$$p(\theta_j | \eta), \quad j = 1, \dots, J. \quad (\text{A.17})$$

Here, θ_j are the group-level parameters that vary by group, while η are the hyperparameters, or population-level parameters, that describe variation across groups.

Level 3 (hyperpriors).

Finally, to represent uncertainty in the population-level distribution, the hyperparameters are assigned a prior density

$$p(\eta). \tag{A.18}$$

The mechanics of partial pooling.

Equations (A.16)–(A.18) implement *partial pooling*: groups with limited or noisy data borrow strength from the shared population distribution, while groups with richer data remain more strongly data-driven.

Intuitively, the hyperparameters η learn what parameter values are typical across groups, and this information regularises estimation of each θ_j [105]. A common special case is the normal group-level model

$$\theta_j \mid \mu, \sigma \sim \mathcal{N}(\mu, \sigma^2), \tag{A.19}$$

where μ is the population mean and σ is the between-group standard deviation. A small σ implies strong pooling because groups are similar, whereas a large σ implies weaker pooling because groups are more distinct.

A.4 Sensitivity to the choice of epidemiological parameters

The baseline simulation chapters use $\mathcal{R}_0 = 3.0$, $\nu = \sqrt{2} \approx 1.41$, and a sustained post-intervention contact level $c^* = 0.3$. This appendix verifies that the qualitative conclusions are not tied to those specific choices by systematically varying each quantity while holding the others fixed. All simulations use the reduced heterogeneous-susceptibility SEIR model in Section 2.2 with the piecewise-linear NPI profile, population size $N = 100,000$, latent period $1/\delta = 5.5$ days, infectious period $1/\gamma = 4$ days, and relative exposed infectiousness $\rho = 0.5$ as given in Table 2.1.

A.4.1 Sensitivity to $\mathcal{R}_0$

Figure A.2 shows daily onset-incidence curves, using the convention in equation (1.10), for $\mathcal{R}_0 \in \{2.5, 3.0, 3.5\}$ with $\nu = \sqrt{2}$ and $c^* = 0.3$. Increasing $\mathcal{R}_0$ raises peak incidence substantially (from approximately 246 to 969 daily cases) and shifts the peak slightly earlier, but the overall epidemic shape – a single wave suppressed by NPIs after day 15, followed by a slow decay – is preserved across the range. The cumulative attack rate over 100 days rises from 6.6% to 22.3%. Thus, while the quantitative scale of the epidemic is sensitive to $\mathcal{R}_0$, the dynamical regime and the role of heterogeneity do not change qualitatively within $\mathcal{R}_0 \in [2.5, 3.5]$.

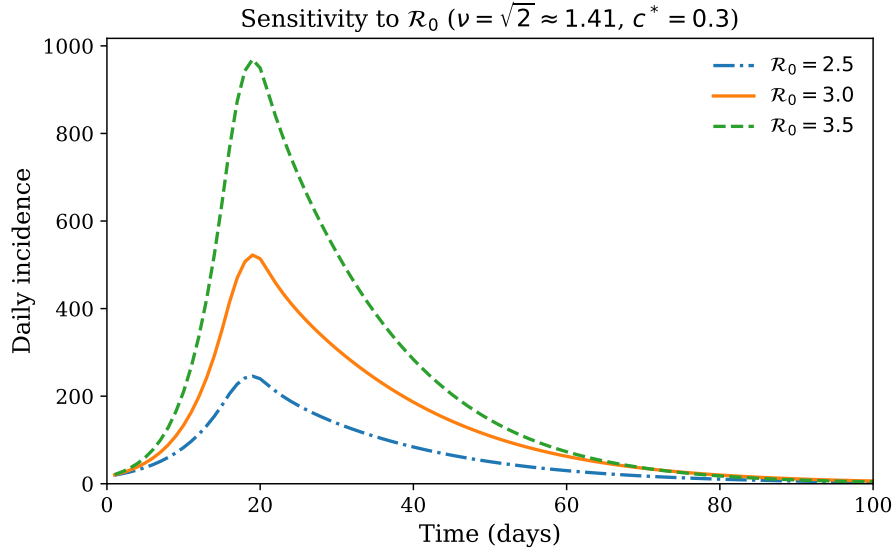


Figure A.2: Sensitivity of daily incidence to the basic reproduction number $\mathcal{R}_0$ under the baseline heterogeneous model ($\nu = \sqrt{2}$, $c^* = 0.3$). Increasing $\mathcal{R}_0$ amplifies peak incidence and total attack size, but the qualitative epidemic trajectory is unchanged.

A.4.2 Joint sensitivity to $\mathcal{R}_0$ and ν

Figure A.3 displays a 3×3 grid combining $\mathcal{R}_0 \in \{2.5, 3.0, 3.5\}$ and $\nu \in \{1.0, \sqrt{2}, 2.0\}$. Two features are apparent. First, for any fixed $\mathcal{R}_0$, increasing ν lowers the peak and accelerates the post-peak decline, reflecting the faster depletion of highly susceptible individuals. Second, the effect of ν on the epidemic tail becomes more pronounced at higher $\mathcal{R}_0$. At $\mathcal{R}_0 = 2.5$, moving from $\nu = 1$ to $\nu = 2$ reduces peak incidence by roughly 13%; at $\mathcal{R}_0 = 3.5$, the same change reduces it by approximately 32%. This interaction does not alter the qualitative conclusions of the thesis, but it does confirm that the practical relevance of correctly specifying ν grows with the underlying transmissibility.

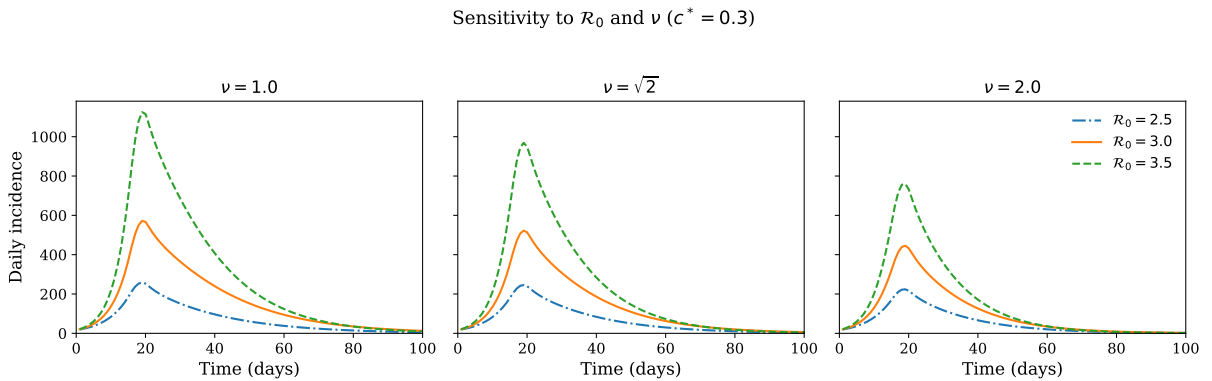


Figure A.3: Joint sensitivity of daily incidence to $\mathcal{R}_0$ and ν (panels), with $c^* = 0.3$. Higher heterogeneity reduces peak incidence and shortens the epidemic tail; this effect is more pronounced at higher $\mathcal{R}_0$.

Appendix B

Robust ν - c^* Correlation Checks across Estimation Methods and NPI Profiles

B.1 Introduction

In Chapter 2, we demonstrated that fitting a heterogeneous-susceptibility SEIR model to a single epidemic wave yields a strong positive correlation between the heterogeneity parameter ν and the intervention floor c^* . This ν - c^* compensation ridge limits practical identifiability: many combinations of ν and c^* can produce similar epidemic trajectories, making it difficult to disentangle population-level heterogeneity from intervention effects using one wave alone.

This appendix stress-tests the robustness of the ν - c^* correlation by systematically varying two key modelling ingredients in a 3×3 factorial design:

1. **Estimation method:** maximum likelihood (MLE), nonlinear least squares (LSQ) and Bayesian inference via Hamiltonian Monte Carlo (MCMC).
2. **NPI functional form:** piecewise-linear (baseline), exponential decay and logistic transition.

The three NPI functional forms, their window-matching derivations, and the parameter values used throughout this appendix are defined in Appendix A.2. The forms are equations (A.12)–(A.14) and the matched parameter values for the simulation window $t_0 = 15$, $t_1 = 20$ days ($w = 5$ days, $c^* = 0.3$) are collected in Table A.1. Figure A.1 in Appendix A.2 shows the three profiles as calibrated to that window. In the robustness analysis, we estimate $(\mathcal{R}_0, \nu, t_0, c^*)$ (while

treating t_1 as fixed) for the linear form in equation (A.12); $(\mathcal{R}_0, \nu, t_{td}, c, \kappa)$ for the exponential form in equation (A.13); and $(\mathcal{R}_0, \nu, t_m, c, k)$ for the logistic form in equation (A.14), comparing estimates of $(\mathcal{R}_0, \nu, c^*)$ across scenarios.

B.2 Simulation Design

For each cell in the 3×3 factorial grid, we perform the following steps:

1. **Data generation.** Simulate 200 independent datasets from the deterministic SEIR model with the NPI profile corresponding to that scenario (linear, exponential, or logistic), using the S4 true parameter values: $\mathcal{R}_0 = 3.0$, $\nu \approx 1.414$, $t_0 = 15$ days, $c^* = 0.30$, plus the profile-specific parameters in Table A.1. Observed daily incidence is generated by Poisson sampling from the expected incidence λ_t .
2. **Model fitting.** Fit each dataset using the same (matched) model and estimation method. For example, datasets generated with the exponential NPI are fitted with the exponential NPI model; datasets generated for the **Bayesian + Logistic** scenario are fitted via MCMC with the logistic $c(t)$ specification.
3. **Analysis.** From the resulting fits (point estimates for MLE and LSQ; posterior samples for Bayesian), compute the Pearson correlation coefficient $\rho(\nu, c^*)$ and summarise parameter recovery. Invalid fits (non-convergence, non-positive-definite Hessian) are excluded which happened in less than 10% of the 200 datasets.

B.3 Estimation Methods

Let y_t denote the observed daily incidence on day t , generated as Poisson draws from the model-implied expected incidence $x_t(\boldsymbol{\theta})$, computed by solving the ODE system (2.1) and extracting the incidence output.

Maximum Likelihood Estimation (MLE)

Parameters are estimated by minimising the negative log-likelihood as in Section 2.5. The log-likelihood is:

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^T \log f(y_t | x_t(\boldsymbol{\theta})) = \sum_{t=1}^T [y_t \log x_t(\boldsymbol{\theta}) - x_t(\boldsymbol{\theta}) - \log(y_t!)], \quad (\text{B.1})$$

where $f(\cdot)$ is the Poisson probability mass function. Optimisation uses the Nelder–Mead simplex algorithm on transformed parameters (log for positive parameters, logit for bounded parameters). The Hessian at the optimum is computed via finite differences to obtain standard errors and correlations.

Least Squares Estimation (LSQ)

Parameters are estimated by minimising the sum of squared errors:

$$\text{SSE}(\boldsymbol{\theta}) = \sum_{t=1}^T [y_t - x_t(\boldsymbol{\theta})]^2. \quad (\text{B.2})$$

While LSQ does not correspond to the formal Poisson likelihood, it is widely used in practice and provides a useful comparison to assess sensitivity to the choice of objective function.

Bayesian estimation.

We implement Hamiltonian Monte Carlo (HMC) sampling in `rstan` with weakly informative priors. The Bayesian model uses the same data model (Poisson likelihood) as the corresponding MLE fit, but targets the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$. Priors are chosen to be weakly informative and centred on plausible ranges:

$$\begin{aligned} \mathcal{R}_0 &\sim \text{LogNormal}(\log 2.5, 0.5), \\ \nu &\sim \text{LogNormal}(0, 1), \\ t_0 &\sim \text{Normal}(15, 5) \text{ [truncated to } t_0 > 0], \\ c^* &\sim \text{Beta}(2, 2). \end{aligned}$$

Profile-specific timing and shape parameters receive similarly weakly informative priors. For each scenario, we run 4 chains of 2000 iterations (1000 warmup, 1000 sampling) with adapt delta $\delta = 0.95$ and maximum tree depth 12, yielding 4000 posterior samples for analysis. Convergence is monitored using $\hat{R}$ and effective sample size diagnostics.

B.4 Results

The analysis consists of nine simulation experiments, corresponding to the nine cells in the 3×3 factorial grid. From the valid MLE and LSQ fits, and from posterior draws in the Bayesian fits, we computed the Pearson correlation coefficient $\rho(\nu, c^*)$. Invalid fits, including non-converged

fits and fits with non-positive-definite Hessians, were excluded from the correlation calculation.

Table B.1 and Figure B.1 summarise the results from all nine scenarios. The central finding is that the positive association between ν and c^* persists across estimation methods and NPI profiles. This is the empirical counterpart of the analytical compensation argument in Section 2.3.2: over a limited part of a single epidemic wave, the data can mainly constrain a composite transmission-reduction factor rather than ν and c^* separately.

B.4.1 Summary of Correlations

Table B.1 reports the Pearson correlation coefficient $\rho(\nu, c^*)$ for each estimation method and NPI profile. All nine cells show positive correlations, ranging from $\rho = 0.651$ for the MLE–logistic fit to $\rho = 0.979$ for the Bayesian–exponential fit, with all but one scenario exceeding 0.89. This consistency shows that higher estimated heterogeneity ν is associated with a higher estimated intervention floor c^* , corresponding to a weaker inferred sustained intervention.

Table B.1: Pearson correlation coefficient $\rho(\nu, c^*)$ across estimation methods and NPI parameterisations in the 3×3 robustness analysis. All correlations are positive, showing that the ν – c^* compensation direction persists across likelihood-based, least-squares, and Bayesian inference, and across exponential, linear, and logistic contact-reduction profiles.

Estimation method	Exponential	Linear	Logistic
MLE	0.891	0.974	0.651
LSQ	0.948	0.976	0.971
Bayesian	0.979	0.944	0.966

B.4.2 Visual Evidence of the Compensation Ridge

Figure B.1 displays the raw estimates (and posterior samples) underlying these correlations. In all nine facets, the strong positive trend is visually evident: the point clouds are elongated along the ν – c^* diagonal, reflecting the fundamental identifiability challenge documented in Chapter 3. The red dashed lines indicate the true parameter values ($\nu \approx 1.414$, $c^* = 0.3$). While estimates cluster around the data-generating values, the pronounced elongation confirms that many combinations of ν and c^* can fit the single-wave incidence data nearly as well as the true values.

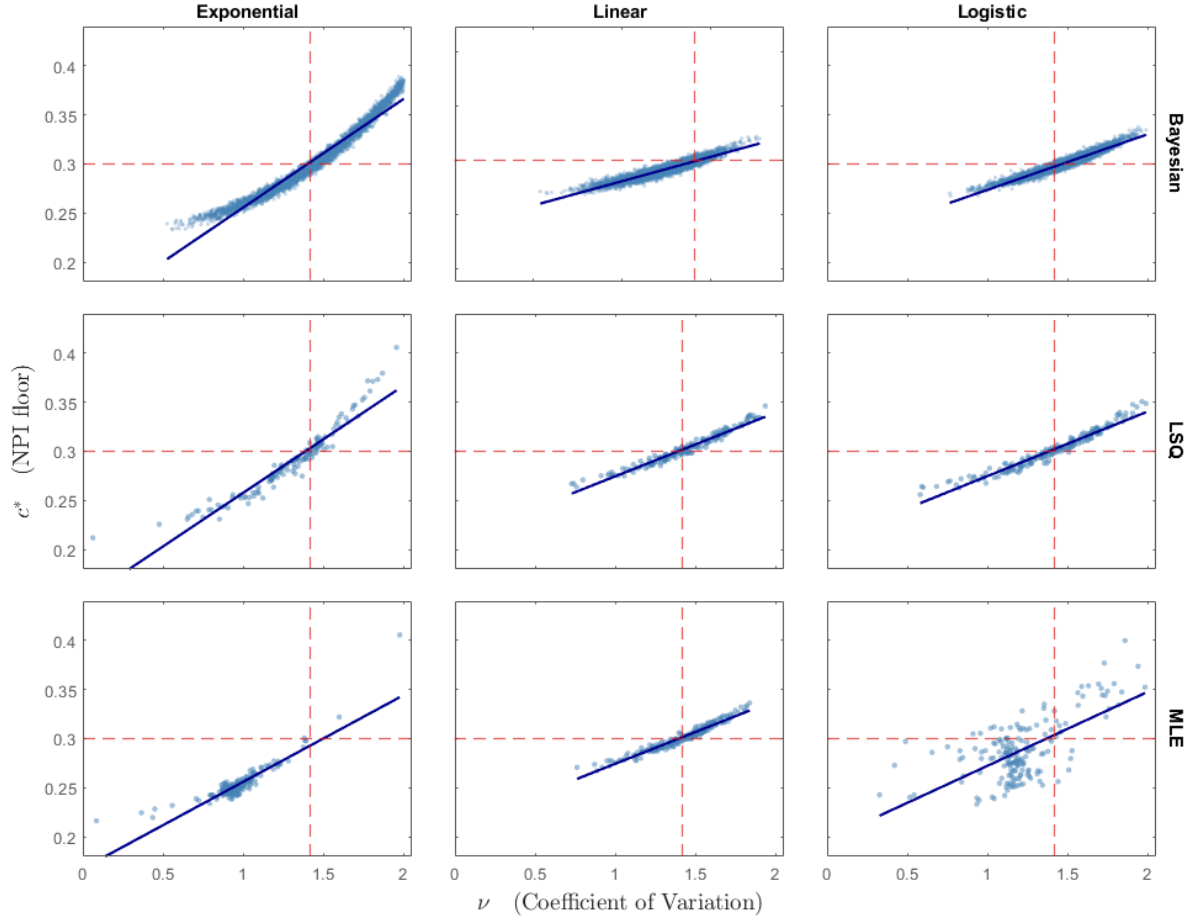


Figure B.1: Scatter plots of ν vs. c^* estimates across all nine scenarios. Raw parameter estimates (MLE and LSQ) or posterior samples (Bayesian) from the 200 simulated datasets in each scenario. The strong positive trend is visible in all facets, confirming the ν - c^* compensation ridge across estimation methods (rows) and NPI profiles (columns). Red dashed lines mark the true values ($\nu \approx 1.414$, $c^* = 0.3$). The elongation along the diagonal indicates that markedly different (ν, c^*) combinations produce similar fits to the observed single-wave data.

B.4.3 Comparison with the analytical compensation argument

The pattern in Figure B.1 is consistent with the analytical compensation argument developed in Section 2.3.2. In that argument, the effective transmission-reduction factor over the intervention-dominated part of the fitted epidemic wave is approximated locally by

$$c_{\text{eff}} \approx c^* \left(\frac{S^*}{N} \right)^{\nu^2},$$

where S^*/N is a representative susceptible fraction over the relevant part of the epidemic curve. Since $0 < S^*/N < 1$, increasing ν decreases the factor $(S^*/N)^{\nu^2}$. To preserve approximately the same c_{eff} , this decrease can be offset by increasing c^* , corresponding to a weaker inferred intervention floor. Conversely, a smaller ν can be offset by a smaller c^* , corresponding to a stronger inferred intervention.

The scatter plots in Figure B.1 show this predicted direction visually: the high-density regions are elongated from lower-left to upper-right in the (ν, c^*) plane. Table B.1 shows that the same positive association is retained when the objective function is changed from Poisson maximum likelihood to least squares and when the inference is carried out in a Bayesian framework. It is also retained when the intervention profile is changed from a piecewise-linear ramp to exponential or logistic forms. Thus, this appendix tests the compensation mechanism already identified analytically in Chapter 2 empirically.

B.5 Conclusion

This appendix has demonstrated that the positive ν - c^* association documented in Chapter 2 is robust to variation in both the estimation method and the functional form of the intervention profile. Across all nine scenarios in the 3×3 factorial design, Pearson correlations remain positive, with $\rho \in [0.651, 0.979]$; see Table B.1. The direction of these correlations agrees with the analytical compensation argument in Section 2.3.2: for a fixed local effective transmission-reduction factor, larger susceptibility heterogeneity is compensated by a larger inferred intervention floor c^* , and vice versa. The robustness analysis therefore supports the interpretation that the ν - c^* ridge is a practical information limitation of single-wave inference rather than an artefact of one fitting method or one assumed NPI curve.

Appendix C

Supplementary Material for Chapters 2 and 3

C.1 Introduction

This appendix provides additional material supporting the identifiability results in Chapter 2 and Chapter 3. Chapter 2 studies the practical limitations of fitting the heterogeneous-susceptibility SEIR model to a single epidemic trajectory, while Chapter 3 studies how joint fitting of two or more related epidemics can reduce those limitations.

The purpose of this appendix is threefold. First, it checks whether the precision gains from joint-epidemic inference remain present when the sustained intervention strength is varied. Second, it provides supplementary likelihood and profile-likelihood diagnostics that complement the main Chapter 3 results. Third, it examines how the correlation structure of the estimation problem changes when the initial conditions are varied, since the multi-epidemic framework in Chapter 3 relies on contrasts between epidemics with different depletion histories.

Chapter 2 showed that, when susceptibility heterogeneity and NPIs are both present, a single epidemic trajectory can generate a strong compensation ridge between ν and c^* : different combinations of heterogeneity and intervention strength can produce very similar incidence curves, making the two effects difficult to separate in practice. The analytical basis for this was developed in Section 2.3.2 through equation (2.8), which shows that the data tend to constrain a composite quantity of the form

$$c_{\text{eff}} \approx c^* (S^*/N)^{\nu^2},$$

rather than ν and c^* separately. Chapter 3 then showed that jointly fitting two epidemics with shared parameters but different initial seeds weakens this ridge and yields substantial gains in

precision and numerical stability.

The analyses in this appendix extend those findings in four directions. Section C.1.1 repeats the single- versus two-epidemic profile-likelihood comparison for three NPI levels, $c^* \in \{0.20, 0.30, 0.40\}$, using the same Large–Small epidemic pair as in Chapter 3. This allows us to assess whether the improvement from joint fitting depends strongly on the overall strength of control. The same section also reports Table C.1, which summarises how the ν – c^* correlation changes with intervention strength for single- and two-epidemic fits.

Section C.1.2 presents bivariate likelihood contours in the (ν, c^*) plane for the two off-baseline intervention levels $c^* = 0.20$ and $c^* = 0.40$. These plots complement the baseline $c^* = 0.30$ likelihood ridge map in Chapter 3. Section C.1.3 then presents representative profile-likelihood curves for $\mathcal{R}_0$ and t_0 , complementing the ν and c^* profile likelihood shown in Chapter 3.

Finally, Section C.2 studies sensitivity to initial conditions and intervention strength by varying the initial infected population $I_i(0) \in \{20, 40, 80, 160, 320, 400\}$ together with $c^* \in \{0.2, 0.3, 0.4\}$. The aim is to see how the main parameter correlations respond to the amount of information available in the early growth phase and to the degree of contrast between epidemic trajectories. This additional check helps connect the main Chapter 3 message—that differences in depletion history improve identifiability, to a broader range of starting conditions.

C.1.1 Profile Likelihood Results Across NPI Strengths

In Section 3.4 of Chapter 3, we showed, for a single NPI strength ($c^* = 0.30$), that combining a “Large” and a “Small” epidemic with very different initial seeds markedly sharpens the profile likelihoods for all four shared parameters $(\mathcal{R}_0, \nu, t_0, c)$. To assess whether this benefit depends on the overall strength of control, we repeated the profile-likelihood analysis for two additional NPI levels ($c^* \in \{0.20, 0.40\}$) alongside the $c^* = 0.30$ results already presented in the chapter, using the same pair of well-separated epidemics with $I_0^{(1)} = 400$ and $I_0^{(2)} = 40$.

For each NPI level we simulated 200 noisy datasets and, for each dataset, computed profile likelihoods and 95% profile-likelihood confidence intervals for both a single-epidemic fit (using only the larger epidemic) and a two-epidemic fit with shared $(\mathcal{R}_0, \nu, t_0, c^*)$.

The reduction in profile-likelihood confidence-interval width was consistent across all three intervention strengths. The largest gains occurred for ν and c^* , the two parameters most directly affected by the ν – c^* compensation ridge. For ν , the two-epidemic fit reduced the interval width by approximately $23.6\times$, $17.2\times$, and $6.3\times$ at $c^* = 0.20, 0.30$, and 0.40 , respectively. For c^* , the corresponding gains were approximately $5.8\times$, $6.5\times$, and $3.4\times$. The remaining parameters also

improved, but more moderately: the gains were approximately $3.3\times$, $2.5\times$, and $2.5\times$ for $\mathcal{R}_0$, and $2.4\times$, $2.2\times$, and $2.4\times$ for t_0 , across the same three NPI levels.

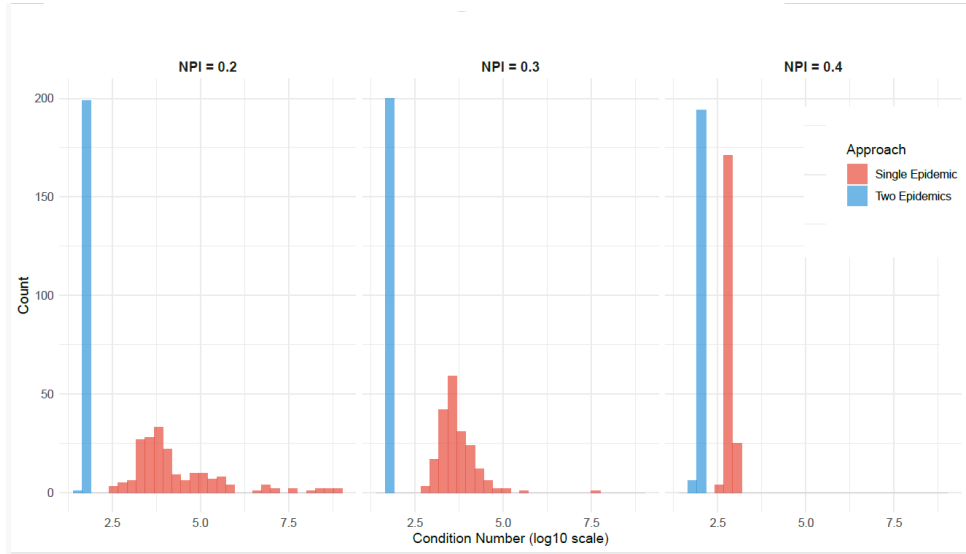


Figure C.1: Distribution of Hessian condition numbers ($\log_{10}$ scale) for single-epidemic versus two-epidemic fits across all three NPI levels. In each panel the single-epidemic condition numbers form a broad, right-skewed distribution spanning several orders of magnitude, whereas the two-epidemic condition numbers are tightly clustered in the tens, indicating better numerical conditioning of the joint likelihood at every intervention strength examined.

Table C.1 summarises the median ν - c^* correlation between profile-likelihood point estimates across 200 simulated datasets, for single- and two-epidemic fits at each of the three NPI levels. Here c^* denotes the contact multiplier (NPI floor), so $c^* = 0.20$ corresponds to an 80% reduction in contacts (strong intervention), $c^* = 0.30$ to a 70% reduction (moderate), and $c^* = 0.40$ to a 60% reduction (weak).

Table C.1: Median ν - c^* Pearson correlation between profile-likelihood point estimates across 200 simulated datasets, for single- and two-epidemic fits at three NPI levels. c^* is the contact multiplier (NPI floor); smaller c^* corresponds to a stronger intervention. The absolute reduction in correlation when moving from one to two epidemics is given in the final column.

NPI floor c^*	Contact reduction	Single epidemic	Two epidemics	Reduction
0.20	80% (strong)	0.93	0.23	0.70
0.30	70% (moderate)	0.96	0.32	0.64
0.40	60% (weak)	0.90	0.68	0.22

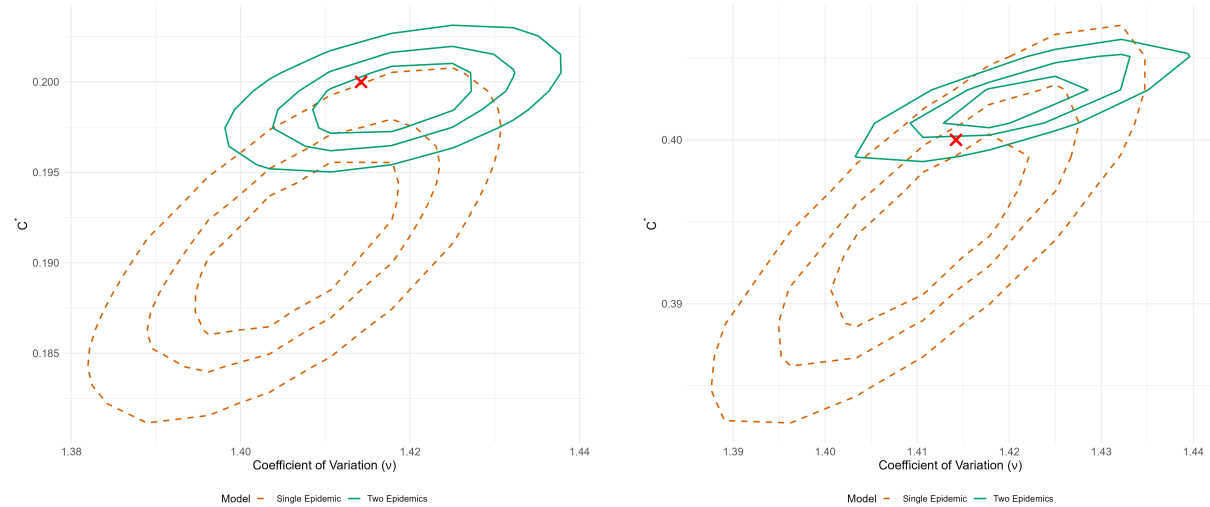
Reading down the single-epidemic column, the ν - c^* correlation is uniformly high at all three NPI levels ($r \geq 0.90$), confirming that the compensation ridge is a robust feature of single-epidemic time series inference that does not weaken as the intervention changes in strength. The two-epidemic results show that joint inference reduces the correlation, but the size of the reduction depends on intervention strength: the drop is largest under strong control ($c^* = 0.20$: $0.93 \rightarrow 0.23$, a reduction of 0.70) and smallest under weak control ($c^* = 0.40$: $0.90 \rightarrow 0.68$, a

reduction of only 0.22), with the moderate baseline case falling between ($c^* = 0.30$: $0.96 \rightarrow 0.32$, a reduction of 0.64).

C.1.2 Bivariate likelihood surfaces across NPI strengths

Figure C.2 compares the (ν, c^*) likelihood contours from the single-epidemic and two-epidemic fits at the two off-baseline intervention levels, $c^* = 0.20$ and $c^* = 0.40$. In each panel, $\mathcal{R}_0$ and t_0 are held fixed, and the contour lines correspond to relative log-likelihood drops of -0.5 , -1 , and -2 from the maximum. The red cross marks the true parameter values. The corresponding baseline case, $c^* = 0.30$, is shown in Chapter 3, Figure 3.4.

The two off-baseline panels show that the qualitative ridge-breaking effect is not restricted to the baseline $c^* = 0.30$ setting. At $c^* = 0.20$, the two-epidemic contours are tight and nearly symmetric, indicating a strong reduction in the ν - c^* ridge. At $c^* = 0.40$, the two-epidemic contours are still narrower than the single-epidemic contours, but show more residual elongation. This agrees with Table C.1, where the residual two-epidemic ν - c^* correlation is smallest under strong intervention and largest under weak intervention.



(a) $c^* = 0.20$, strong intervention.

(b) $c^* = 0.40$, weak intervention.

Figure C.2: Single-epidemic contours versus two-epidemic contours in the (ν, c^*) plane for the two off-baseline NPI levels. The single-epidemic contours are elongated along the compensation ridge, while the two-epidemic contours are more concentrated. The ridge-breaking effect is strongest under $c^* = 0.20$ and weaker, though still visible, under $c^* = 0.40$. The baseline $c^* = 0.30$ case is shown in the main text, Figure 3.4.

C.1.3 Profile likelihood curves for the remaining shared parameters

Figure 3.5 in Chapter 3 shows the profile likelihoods for the two parameters most directly involved in the ν - c^* compensation ridge. Figure C.3 shows the corresponding profile likelihoods for the

remaining shared parameters, $\mathcal{R}_0$ and t_0 , using the same representative baseline $c^* = 0.30$ setting. Together, these profiles show that the two-epidemic design sharpens both the ridge parameters and the remaining shared parameters.

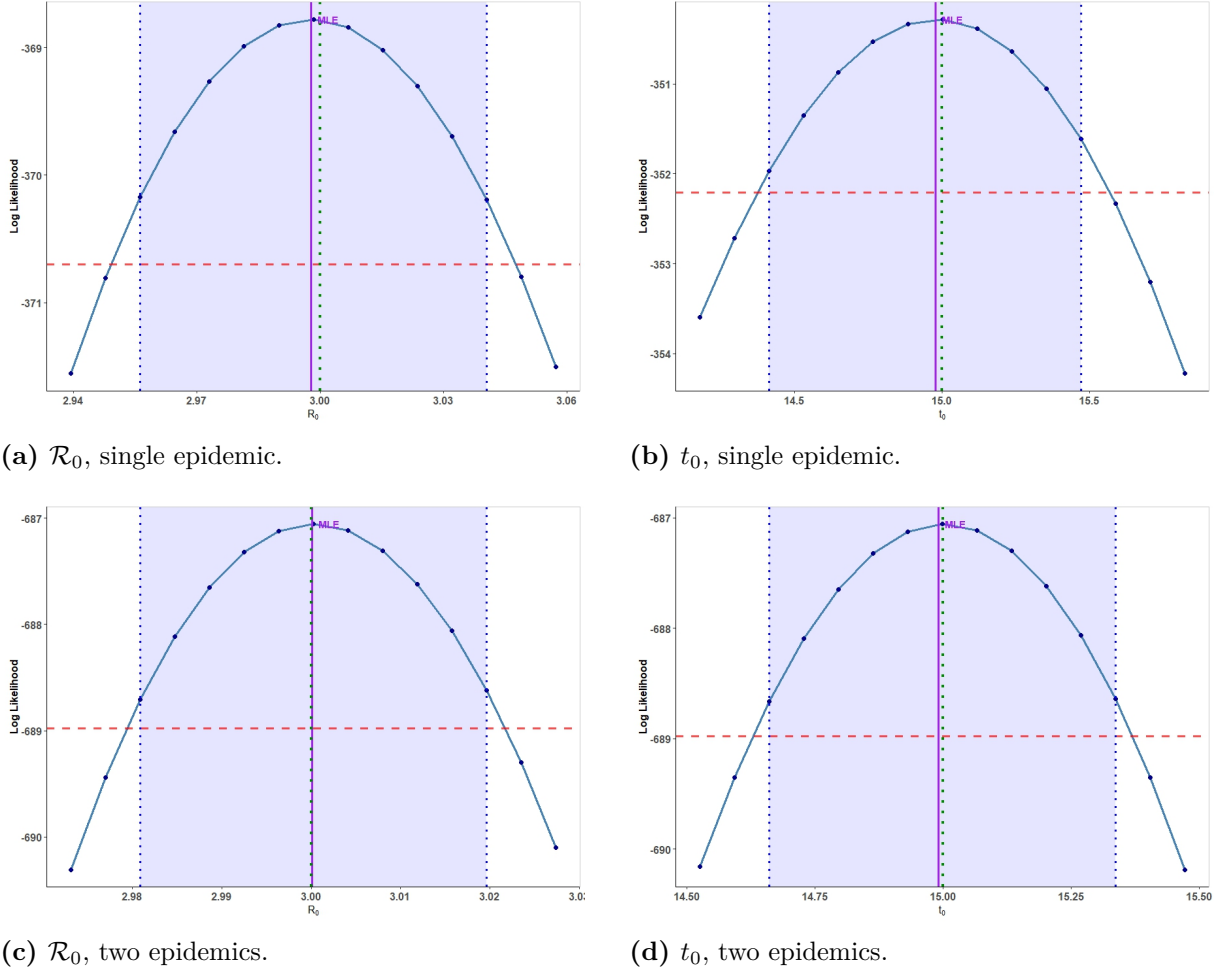


Figure C.3: Profile likelihood curves for the basic reproduction number $\mathcal{R}_0$ and behavioural-change timing t_0 in a representative baseline dataset with $c^* = 0.30$. Horizontal dashed red lines indicate the 95% profile-likelihood threshold; green dotted vertical lines mark the data-generating values, and purple vertical lines mark the MLEs. The two-epidemic profiles are sharper, consistent with the shorter 95% confidence-interval widths reported in Table 3.1.

C.2 Sensitivity to initial conditions and intervention strength

This section examines whether the correlation structure observed in Chapters 2 and 3 depends materially on the initial seed size of the epidemic and on the strength of the intervention. The motivation is that the two-epidemic framework gains information from contrasts in early growth and depletion history, so it is useful to check how these features change when the initial number infected is varied.

We consider initial infected populations

$$I_i(0) \in \{20, 40, 80, 160, 320, 400\},$$

together with three sustained intervention levels

$$c \in \{0.2, 0.3, 0.4\}.$$

For each combination, we generated 200 synthetic datasets from the heterogeneous SEIR model using the fixed parameters in Table 2.1, and then estimated the model parameters as in Chapter 3. The aim is to study how the main pairwise correlations between the estimated parameters change as the epidemic is seeded earlier or later in practical terms.

Figures C.4, C.5, and C.6 summarise these correlation distributions for $c^* = 0.4$, $c^* = 0.3$, and $c^* = 0.2$, respectively. In each figure, the blue marker indicates the baseline initial condition used in the main chapter, namely $I_i(0) = 40$.

Three patterns are clear. First, the correlation between $\mathcal{R}_0$ and ν generally increases with the initial infected population. This suggests that as the epidemic is seeded with a larger initial number infected, less information is available from the very early growth phase to separate intrinsic transmission strength from the effect of selective depletion.

Second, the correlation between ν and c^* remains high across most of the initial conditions considered. Hence, varying the seed size does not remove the basic heterogeneity–intervention trade-off already identified in the single-epidemic analysis. This is consistent with the main message of Chapters 2 and 3: the ν – c^* compensation is a robust feature of single-epidemic series inference rather than a consequence of one particular starting condition.

Third, the relationship between $\mathcal{R}_0$ and t_0 is less stable. For smaller initial seeds the correlation is mostly negative, whereas for larger seeds it can become positive. This change in sign indicates that the interaction between epidemic growth rate and intervention timing depends on the amount of early-phase information available and on how that information is redistributed across the other fitted parameters. For this reason, the $\mathcal{R}_0$ – t_0 dependence is best interpreted as a secondary feature of the likelihood geometry, rather than the main source of non-identifiability.

In summary, these results show that initial conditions do influence the detailed correlation structure of the estimation problem, but they do not alter the central conclusion. The dominant single-epidemic difficulty remains the confounding between susceptibility heterogeneity and intervention strength, while differences in initial conditions mainly affect how strongly the other

parameters are able to compensate around that ridge.

Figures C.4, C.5, and C.6 show the distribution of parameter correlations across initial conditions for both single and two epidemic approaches.

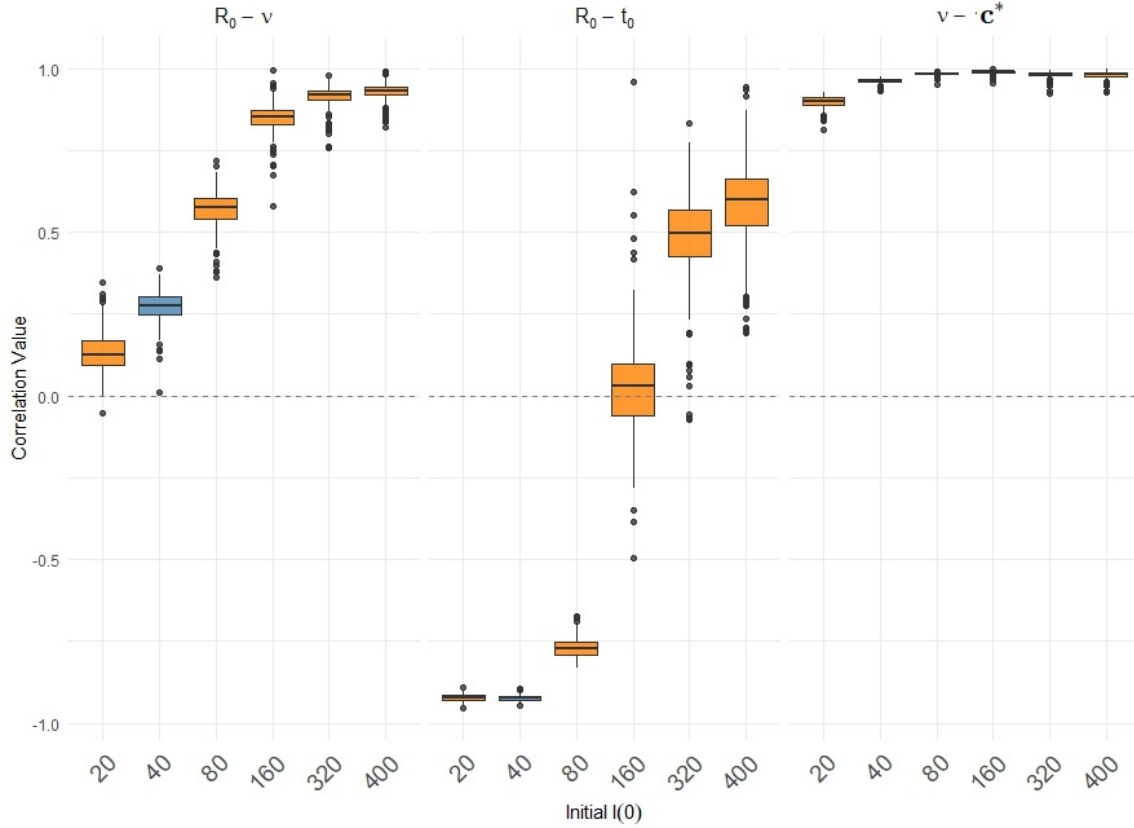


Figure C.4: Correlation distributions in terms of initial condition ($I_i(0) \in \{20, 40, 80, 160, 320, 400\}$) for intervention parameter $c^* = 0.4$. Cases with initial conditions used in the baseline analysis ($I_i(0) = 40$) are marked in blue.

From the point of view of Chapter 3, these results suggest that effective ridge-breaking requires a pair of epidemics with a clear contrast in early-growth and depletion history. The baseline Large–Small design used in Chapter 3 was chosen to provide such a contrast. While the present sensitivity analysis does not identify a precise cutoff for the minimum separation required, it indicates that the gain from joint inference is driven by using epidemics that are sufficiently different in stage to constrain the compensating ν – c^* direction.

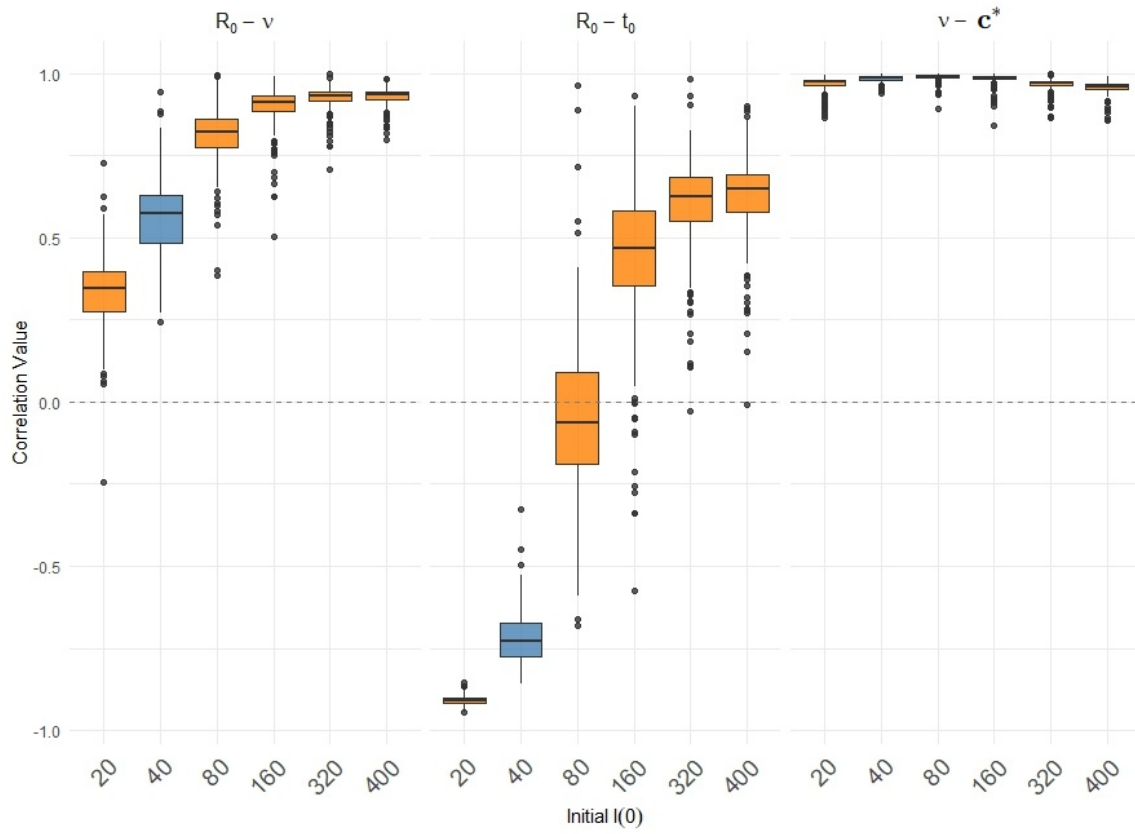


Figure C.5: Correlation distributions in terms of initial condition ($I_i(0) \in \{40, 80, 160, 320, 400\}$) for intervention parameter $c^* = 0.3$. Cases with initial conditions used in the baseline analysis ($I_i(0) = 40$) are marked in blue.

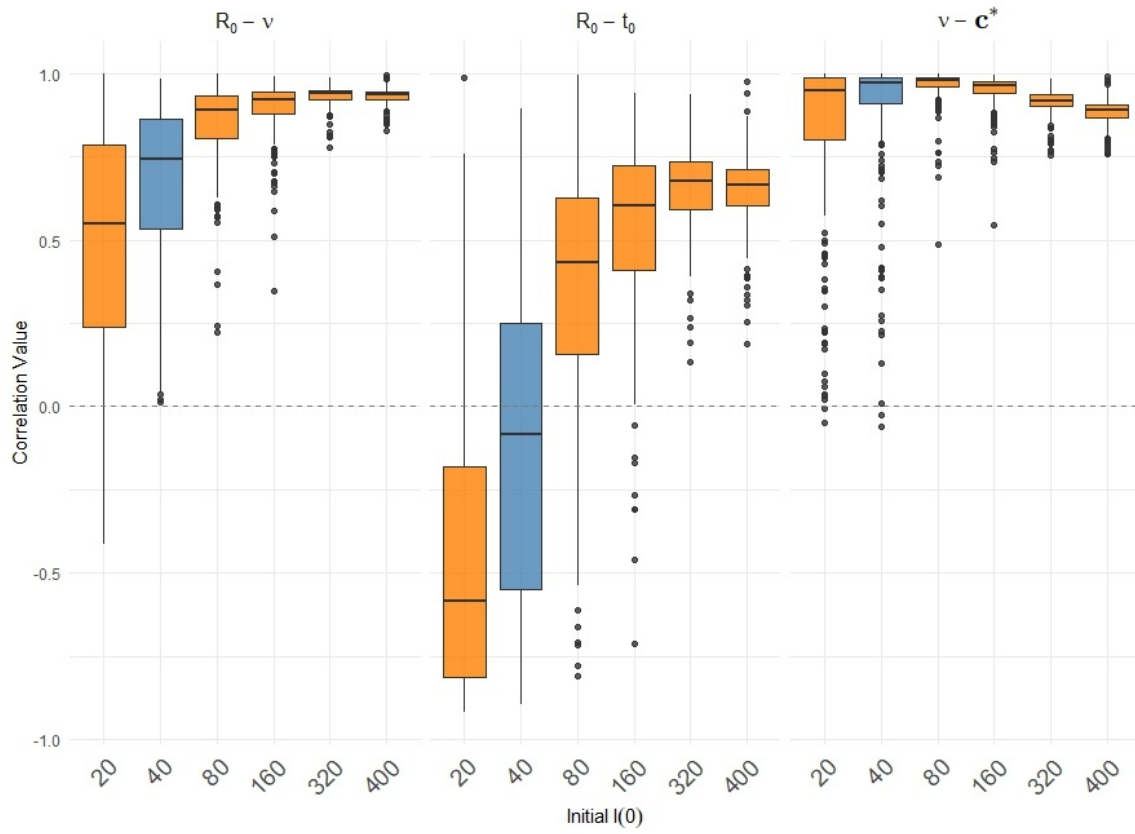


Figure C.6: Correlation distributions in terms of initial condition ($I_i(0) \in \{20, 40, 80, 160, 320, 400\}$) for intervention parameter $c^* = 0.2$. Cases with initial conditions used in the baseline analysis ($I_i(0) = 40$) are marked in blue.

Appendix D

Higher-Dimensional Parameter Spaces via Bayesian Inference

The profile-likelihood analysis in Chapter 3 relied on a relatively low-dimensional parametrisation with a single set of shared parameters $(\mathcal{R}_0, \nu, t_0, c^*)$ across epidemics. This was sufficient to show that two epidemics are enough to alleviate the ν - c^* compensation ridge and to quantify the diminishing returns from adding a third epidemic. However, in more realistic settings we may wish to estimate additional parameters, such as epidemic-specific behavioural onset times, different NPI strengths, or region-specific transmission rates.

Relaxing these assumptions increases the dimensionality of the parameter space and strengthens potential posterior correlations, making a Bayesian inference a more suitable method to use via Markov chain Monte Carlo (MCMC) simulations. Thus, we extend the multi-epidemic likelihood of Section 3.3 to a Bayesian setting.

D.1 Scenarios, Simulations, Model and Priors

We use the reduced heterogeneous-susceptibility SEIR model in equation (2.1). For each epidemic, daily incidence is modelled using the same observation model as in Chapter 3.

Scenarios

We consider three scenarios for two countries with similar population structures where heterogeneity and interventions are the same.

- (i) **Baseline case (4-parameter model):** This replicates the baseline setting of Chapter 3, Section 3.4: two countries with different initial conditions and a fully shared parameter

vector to estimate: $\boldsymbol{\theta} = (\mathcal{R}_0, \nu, t_0, c^*)$;

(ii) **Scenario A (5-parameter model):** permits the onset of behavioural changes to start at different times (but the lockdown occurs at the same time). The epidemics share $(\mathcal{R}_0, \nu, c^*)$, with the distinct $t_0^{(1)}$ and $t_0^{(2)}$ to estimate: $\boldsymbol{\theta}_A = (\mathcal{R}_0, \nu, t_0^{(1)}, t_0^{(2)}, c^*)$;

(iii) **Scenario B (6-parameter model):** further permits growth to be different in the two areas. The epidemics have epidemic-specific reproduction numbers and onset times, $(\mathcal{R}_0^{(1)}, \mathcal{R}_0^{(2)}, t_0^{(1)}, t_0^{(2)})$, while keeping ν and c^* common, hence we estimate: $\boldsymbol{\theta}_B = (\mathcal{R}_0^{(1)}, \mathcal{R}_0^{(2)}, \nu, t_0^{(1)}, t_0^{(2)}, c^*)$.

Simulations

In all three cases, we simulate the synthetic datasets using the same baseline two-epidemic “Large” and “Small” design as in Section 3.3.4, Population size $N = 100,000$, observation window 100 days. Initial conditions: Large epidemic ($E_0 = 1000$, $I_0 = 400$), Small epidemic ($E_0 = 50$, $I_0 = 40$), with true values $\mathcal{R}_0 = 3.0$, $\nu = 1.414$, and $c = 0.3$, $t_1 = 20$ and the NPI profile follows the same piecewise linear profile in equation (2.3).

Model

The two-epidemic analysis in each scenario assumes conditional independence between epidemic trajectories given $\boldsymbol{\theta}$, so the joint likelihood factorises as in equation (3.3).

The joint posterior target for a shared parameter vector $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta} \mid \mathbf{y}_1, \mathbf{y}_2) \propto L_1(\boldsymbol{\theta} \mid \mathbf{y}_1) L_2(\boldsymbol{\theta} \mid \mathbf{y}_2) p(\boldsymbol{\theta}), \quad (\text{D.1})$$

implemented using Hamiltonian Monte Carlo (HMC) via rStan.

Priors

Priors are chosen to be weakly informative within plausible range with uniform distribution spanning over the true simulation values for all the scenarios as:

$$\mathcal{R}_0^{(1,2)} \sim \text{Uniform}(1, 6) \quad (\text{D.2})$$

$$\nu \sim \text{Uniform}(0.5, 3) \quad (\text{D.3})$$

$$t_0^{(1,2)} \sim \text{Uniform}(5, 30) \quad (\text{D.4})$$

$$c^* \sim \text{Uniform}(0.1, 0.9) \quad (\text{D.5})$$

D.2 Baseline Case: 4-Parameter Model

We start by replicating the main single- versus two-epidemic results of Chapter 3, estimating the four-parameter vector: $\theta = (\mathcal{R}_0, \nu, t_0, c^*)$.

D.2.1 Posterior Summaries and precision Gains

Table D.1 compares posterior medians and central 95% credible intervals (CrI) summary for 200 synthetic datasets of both single-epidemic fit and the two-epidemic joint fit. Figure D.1 shows the

Table D.1: Posterior summaries for the shared parameters under single- and two-epidemic Bayesian fits (four-parameter model; NPI $c^* = 0.30$). For each parameter, we report the posterior median and the width of the central 95% credible interval (CrI).

Parameter	Single Epidemic		Two Epidemics	
	Median [2.5%, 97.5%]	CrI Width	Median [2.5%, 97.5%]	CrI Width
$\mathcal{R}_0$	2.979 [2.869, 3.084]	0.215	2.987 [2.959, 3.014]	0.055
ν	1.408 [1.149, 1.594]	0.445	1.425 [1.392, 1.457]	0.065
t_0	15.12 [14.12, 16.08]	1.96	15.18 [14.74, 15.62]	0.88
c^*	0.295 [0.257, 0.332]	0.075	0.298 [0.294, 0.303]	0.009

combined posterior densities for the single- and two-epidemic Bayesian fits. The two-epidemic posterior is visibly more concentrated around the true simulation values. Quantitatively, moving from one to two epidemics reduces the 95% credible-interval widths by factors of approximately $3.9\times$ for $\mathcal{R}_0$, $6.9\times$ for ν , $2.2\times$ for t_0 , and $8.9\times$ for c^* . The largest gains are therefore obtained for ν and c^* , the two parameters most directly affected by the ν - c^* compensation ridge in single-epidemic inference.

D.2.2 Fit quality and Diagnostics: Baseline

Figure D.2 provides posterior residual diagnostics for the two-epidemic model. The residuals are evenly distributed around zero without any discernible temporal patterning, indicating a very good model fit. The Root Mean Square Error (RMSE) values are relatively low, measuring 10.70 for Epidemic 1 and 8.76 for Epidemic 2.

The fitted vs observed trajectory is shown in Figure D.3. The fits shows a very good agreements of the model with the synthetic data. For each epidemic trajectory, points represent the observed daily cases, solid lines denote the posterior predictive median, and the shaded

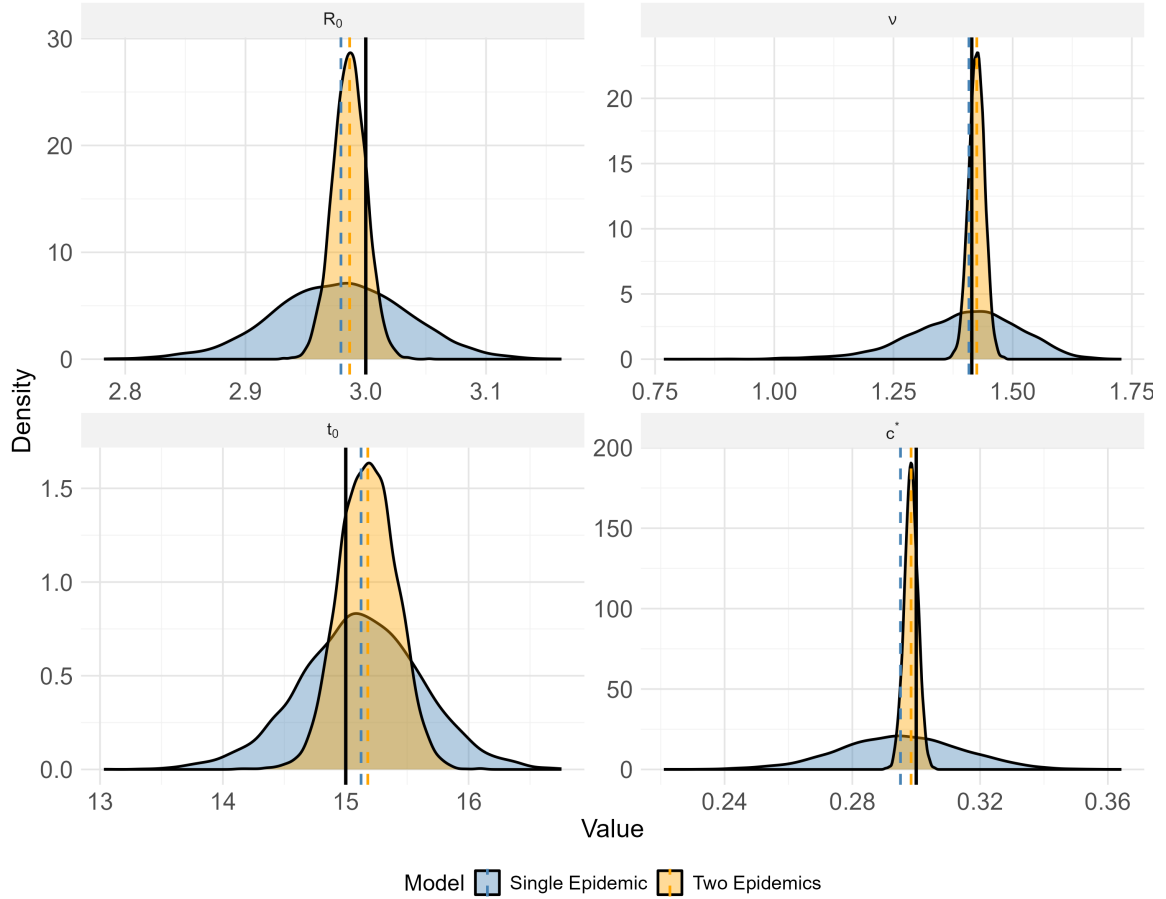


Figure D.1: Posterior density comparison for the four shared parameters under single- and two-epidemic Bayesian fits. Blue distributions show single-epidemic posteriors; orange distributions show two-epidemic joint posteriors. Solid red vertical lines mark true values; dashed lines mark posterior medians. The two-epidemic model narrows posterior uncertainty across all parameters. Transitioning from one to two epidemics narrows the 95% credible interval widths by approximately $3.9\times$ for R_0 , $6.9\times$ for ν , $2.2\times$ for t_0 , and $8.9\times$ for c^* , so that the two-epidemic posterior concentrates tightly around the true simulation values.

ribbons indicate the central 95% predictive intervals. The vertical dashed red line marks the true timing of behavioural change, t_0 . Figure D.4 shows trace plots for four chains of both models: single- and two-epidemic fits. The two-epidemic traces are visibly tighter due to the stronger posterior concentration, but both fits show stable stationarity and good mixing, with no evident multimodality at the scale shown.

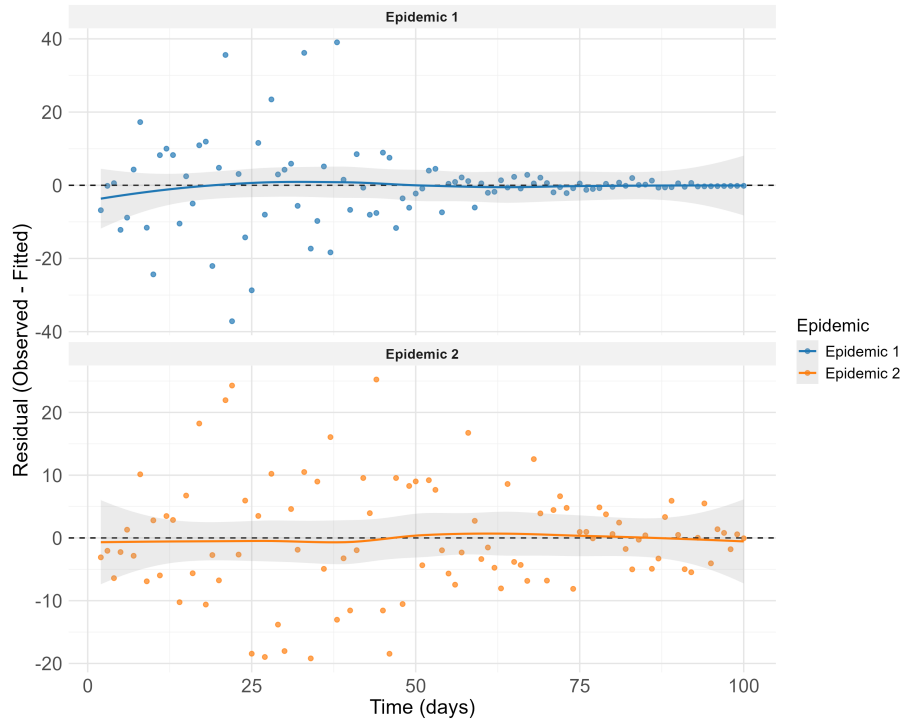


Figure D.2: Posterior residual diagnostics for the two-epidemic Bayesian fit. Residuals are calculated as the observed daily incidence minus the fitted posterior median. The solid lines represent a smoothed trend of the residuals, with shaded regions indicating the 95% uncertainty intervals. The symmetric distribution of residuals around the zero line (dashed) with no obvious pattern suggests an absence of systematic bias. RMSE values are 10.70 for Epidemic 1 (top) and 8.76 for Epidemic 2 (bottom).

D.2.3 Correlations and the ν - c^* ridge: Baseline

A noticeable difference is evident when comparing the Bayesian posterior correlation structure to the likelihood-based (Hessian) results presented in Table 3.3 of Chapter 3. While the MLE-based Hessian analysis indicates a mix of positive and negative correlations for the single-epidemic model (notably a strong negative correlation of approximately -0.70 between $\mathcal{R}_0$ and t_0), the selected single-epidemic MCMC posterior correlations in Table D.2 are positive for the main parameter pairs discussed here.

In contrast, with the two-epidemic joint MCMC model (Table D.2), the strong ν - c^* dependence is substantially moderated, dropping from 0.97 to 0.32. This is similar to the Hessian-based correlation reduction observed in the main text. Furthermore, negative correlations re-emerge

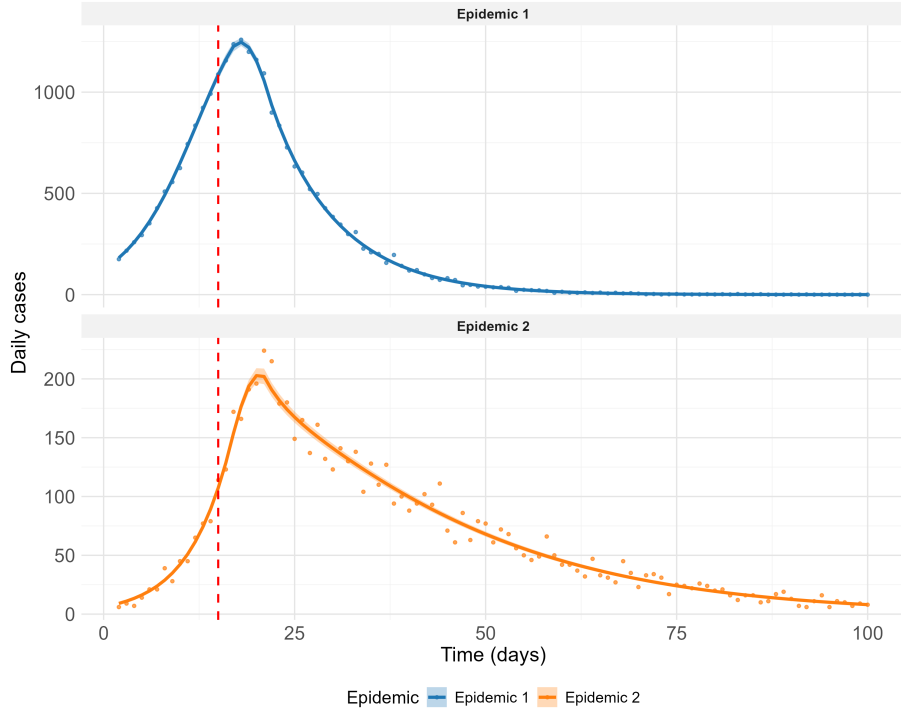


Figure D.3: Fitted versus observed daily incidence for the two-epidemic Bayesian model. The scatter points represent the daily observed cases for Epidemic 1 (top) and Epidemic 2 (bottom). The solid lines denote the fitted posterior median trajectories, while the shaded ribbons represent the central 95% posterior predictive intervals. The vertical dashed red line indicates the true onset time of behavioural change (t_0). Fitting both trajectories simultaneously under the shared parameter vector $(\mathcal{R}_0, \nu, t_0, c^*)$ yields a tight predictive ribbon and concentrated posterior.

between $\mathcal{R}_0$ and t_0 (-0.64), as well as between $\mathcal{R}_0$ and c^* (-0.52). In short, both MLE and Bayesian models, demonstrate consistent identifiability gains under multi-epidemic inference, with the structural trade-off between baseline transmission ($\mathcal{R}_0$) and intervention timing (t_0) persisting even in the Bayesian setting.

Table D.2: Selected posterior correlations for the Bayesian 4-parameter model under single-epidemic and two-epidemic fitting. The table reports the parameter pairs discussed in the text. Values are Pearson correlations computed from the posterior draws.

Parameter pair	Single epidemic	Two epidemics
$\mathcal{R}_0-t_0$	0.69	-0.64
$\mathcal{R}_0-c^*$	0.93	-0.52
$\nu-c^*$	0.97	0.32

D.3 Scenario A:5-parameter Model

In Scenario A we retain shared $(\mathcal{R}_0, \nu, c^*)$ across the two epidemics but allow the onset of behavioural change to differ between epidemics via $t_0^{(1)}$ and $t_0^{(2)}$. The estimated parameter vector is $\theta_A = (\mathcal{R}_0, \nu, t_0^{(1)}, t_0^{(2)}, c)$,

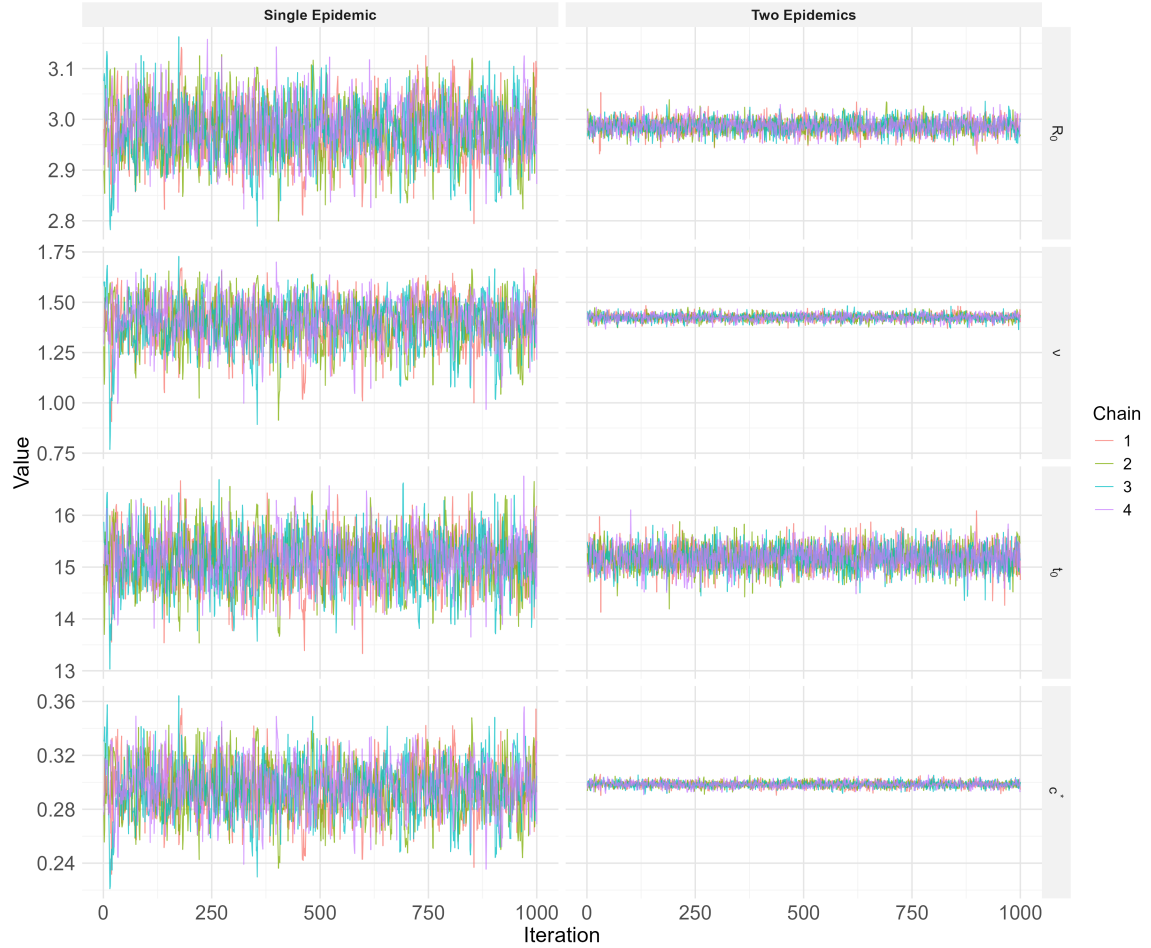


Figure D.4: MCMC trace plots comparing single- and two-epidemic four-parameter fits. **Left column (Single Epidemic):** each of the parameters ($\mathcal{R}_0, \nu, t_0, c^*$) shows wide exploration ranges but good mixing, with each parameter converging near the true value but with wider uncertainty than the corresponding two-epidemic fit. **Right column (Two Epidemics):** all parameters show excellent mixing and tight convergence around their true values. ν and c^* display highly stable traces centred around 1.414 and 0.3, with narrower convergence as expected under joint inference. $\mathcal{R}_0$ and t_0 show a similar pattern.

with data generated at $\mathcal{R}_0 = 3.0$, $\nu = 1.414$, $c^* = 0.3$, $t_0^{(1)} = 12$ and $t_0^{(2)} = 18$, with a fixed $t_1 = 20$. Thus population 1 begins to change behaviour six days earlier than population 2, while both ultimately converge to the same intervention floor $c^* = 0.3$.

Table D.3: Posterior summaries for the two-epidemic model with distinct behavioural onset times $t_0^{(1)}$ and $t_0^{(2)}$ (cases-only likelihood). Entries are posterior medians with posterior SD and central 95% CrIs. “Width” is the CrI width.

Parameter	True	Median	SD	2.5%	97.5%	Width
$\mathcal{R}_0$	3.000	3.020	0.018	2.987	3.057	0.070
ν	1.414	1.394	0.030	1.335	1.454	0.118
$t_0^{(1)}$	12.00	11.46	0.35	10.80	12.16	1.35
$t_0^{(2)}$	18.00	17.91	0.38	17.21	18.64	1.42
c^*	0.3000	0.2963	0.0029	0.2907	0.3024	0.0117

D.3.1 Posterior summaries: Scenario A

Table D.3 reports posterior medians, posterior SDs, and central 95% credible intervals (CrIs) while Figure D.5 shows marginal posterior densities with the true values (red vertical lines) and posterior medians (blue dashed lines). All parameters are recovered with small bias relative to the actual true values: the posterior median for $\mathcal{R}_0$ is 3.020 with standard deviation 0.018 (95% CrI width 0.070), while the coefficient of variation has median $\nu = 1.394$ (SD 0.030, CrI width 0.118). The intervention strength is similarly well constrained, with posterior mean $c = 0.296$ (SD 0.003) and a very narrow 95% interval of width 0.012. These results indicate that sharing $(\mathcal{R}_0, \nu, c^*)$ across epidemics continues to stabilise inference even when the timing of behaviour change is allowed to vary.

The onset (behavioural change) times themselves are estimated with moderate uncertainty but remain well separated. The posterior means are $t_0^{(1)} = 11.46$ (SD 0.350, CrI width 1.35 days) and $t_0^{(2)} = 17.91$ (SD 0.38, CrI width 1.42 days).

D.3.2 Fit quality and diagnostics: Scenario A

Model fit quality and residual diagnostics are visualised in Figure D.7 and Figure D.6, respectively. The fitted trajectories demonstrate agreement with the observed data, while the residuals shows no systematic temporal bias. MCMC convergence is robust across all four chains, as evidenced by the stable and well-mixed trace plots (Figure D.8).

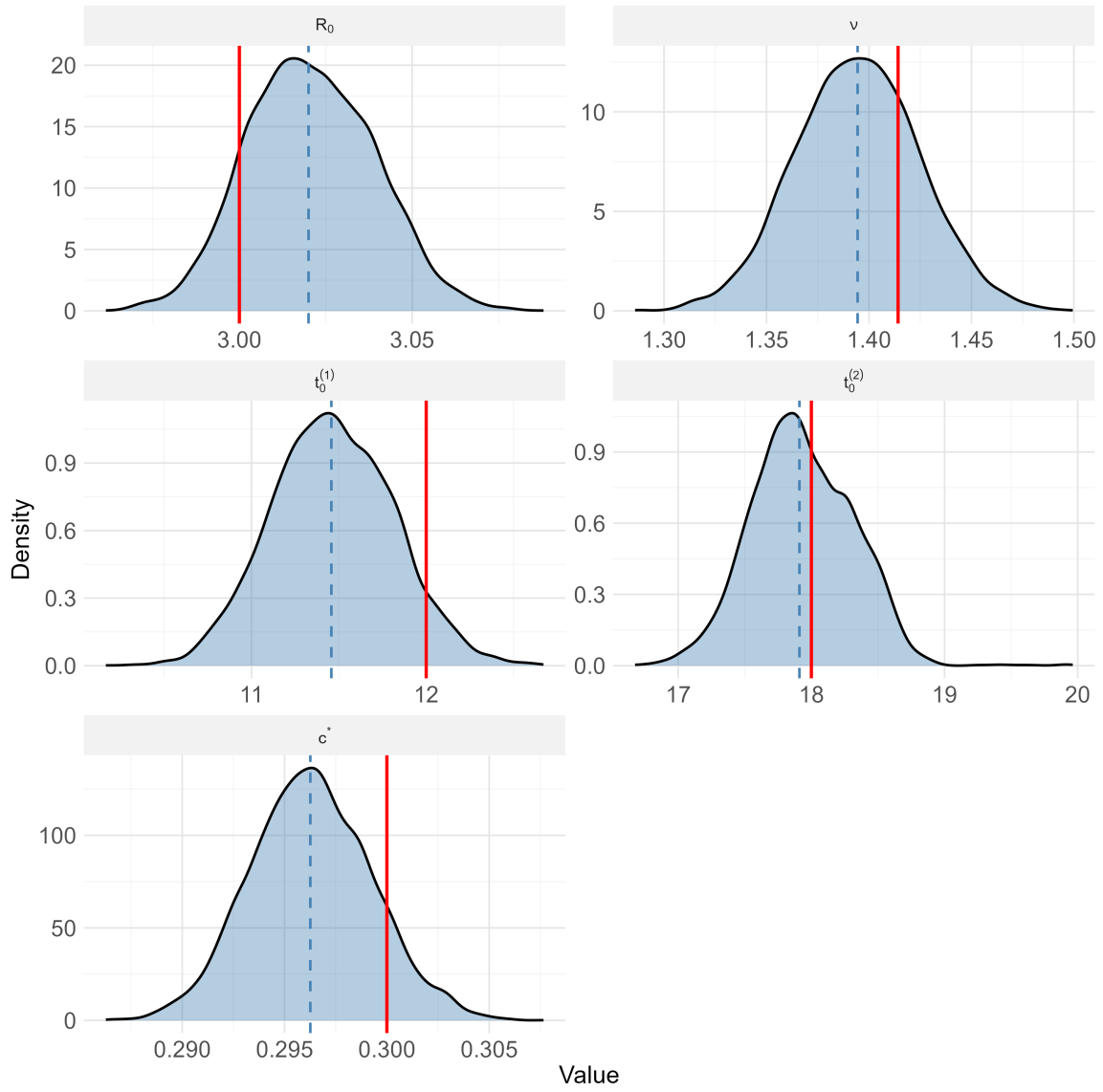


Figure D.5: Marginal posterior densities for the 5-parameter model (Scenario A). Solid red vertical lines mark the true data-generating values, while dashed blue lines indicate the posterior medians. The model successfully recovers the shared parameters alongside the epidemic-specific onset times. Although the true parameters fall within the 95% credible interval, only $t_0^{(2)}$ has its point estimate centred on the true value, followed closely by ν . The earlier onset time $t_0^{(1)}$ shows a more visible shift from the true value.

D.3.3 Correlations: scenario A

Table D.4 clarifies how the two-epidemic model accommodates the extra degree of freedom. The correlation between ν and c^* is 0.71 is slightly less than that for single epidemic 4-parameter model (0.97) in Table D.2 and much higher than that of its counter-part two-epidemic model which is 0.32. This is the price for the additional degree of freedom of extra parameter. Again, we note that $(t_0^{(1)}, \nu)$ as well as $(t_0^{(1)}, c^*)$ have nearly equivalent correlation of about 0.65 while the persistent (from Chapter 3) $R_0 - t_0$ strong negative correlation (-0.82) occurred with $t_0^{(2)}$ while with $t_0^{(1)}$ is lower (-0.37). As a whole, however, the joint posterior remains single-peaked and

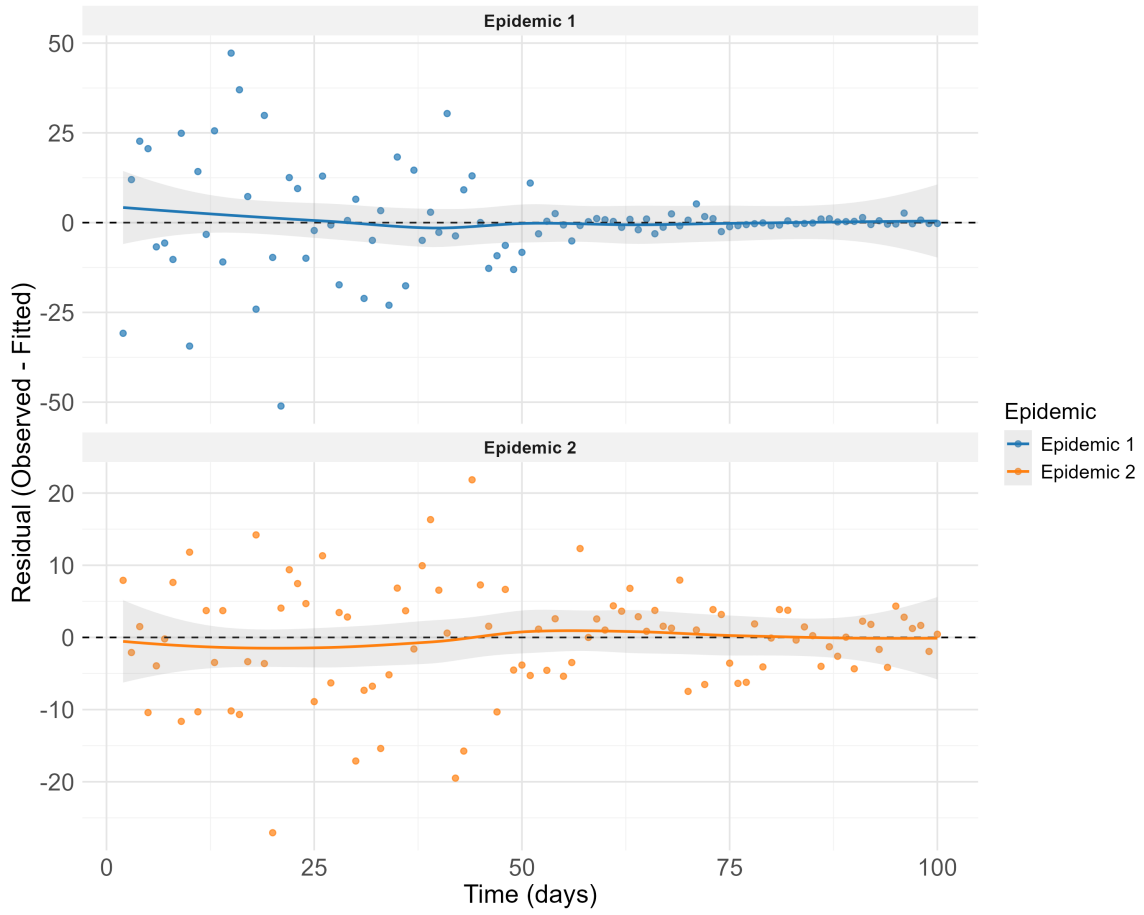


Figure D.6: Residuals versus time for the 5-parameter model (Scenario A). Residuals calculated as observed incidence minus the fitted posterior median. The solid lines represent the smoothed residual trends with 95% uncertainty bands. The symmetric distribution around zero indicates an adequate model fit, with RMSE values of 13.32 for Epidemic 1 and 7.50 for Epidemic 2.

well behaved, and the two-epidemic model continues to provide practically identifiable estimates in this modestly higher-dimensional setting.

Table D.4: Selected posterior correlations for the 5-parameter model (Scenario A). The table reports the parameter pairs discussed in the text, with emphasis on the partial re-emergence of ν - c^* dependence and the $\mathcal{R}_0$ - t_0 timing trade-off.

Parameter pair	Posterior correlation
$\mathcal{R}_0$ - $t_0^{(1)}$	-0.37
$\mathcal{R}_0$ - $t_0^{(2)}$	-0.82
ν - $t_0^{(1)}$	0.66
ν - c^*	0.71
$t_0^{(1)}$ - c^*	0.65

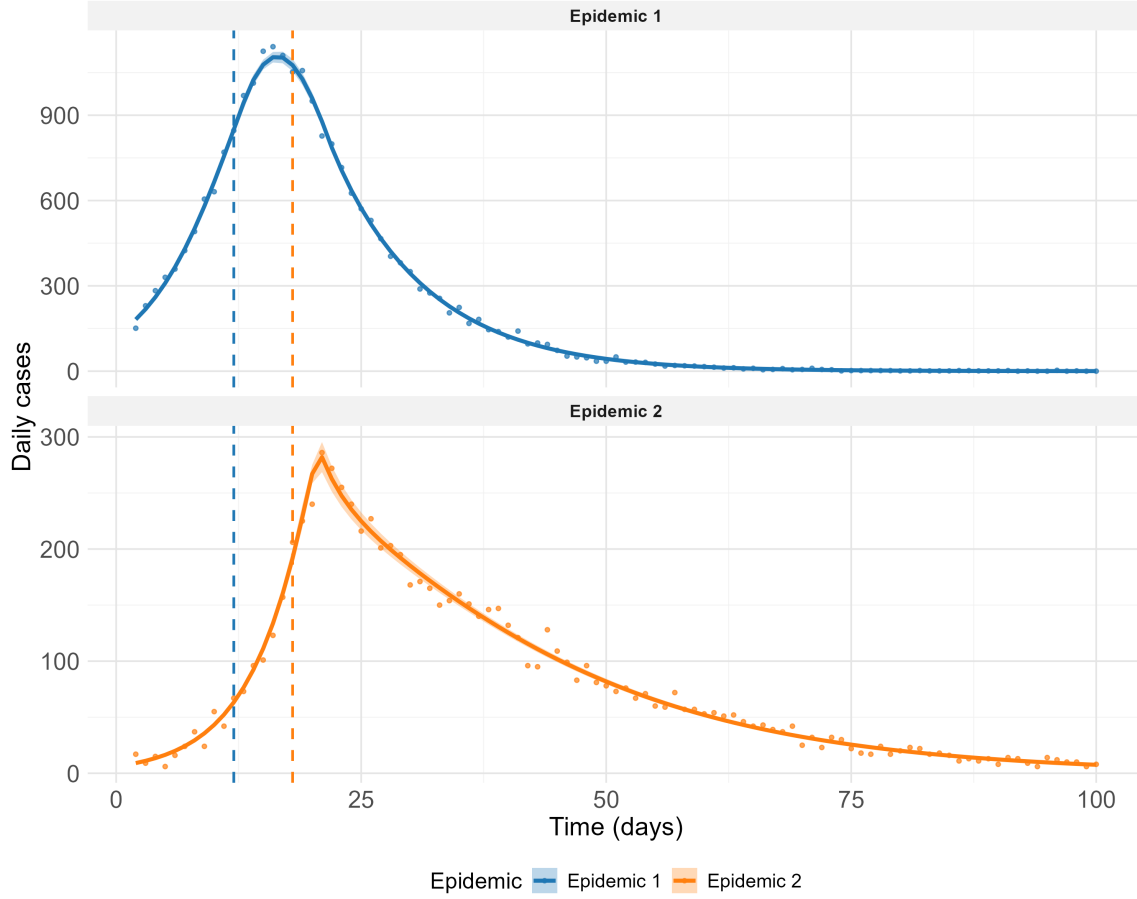


Figure D.7: Fitted versus observed daily incidence for the two epidemics under the 5-parameter model. Scatter points represent the observed daily cases for Epidemic 1 (top) and Epidemic 2 (bottom). Solid lines trace the fitted posterior medians, surrounded by 95% posterior predictive ribbons. The vertical dashed lines mark the true, epidemic-specific onset times of behavioural change ($t_0^{(1)}$ and $t_0^{(2)}$).

D.4 Scenario B:6-parameter Model

Scenario B further relaxes cross-epidemic sharing by allowing each epidemic to have its own reproduction number and onset time while keeping (ν, c^*) shared:

$$\theta_B = (\mathcal{R}_0^{(1)}, \mathcal{R}_0^{(2)}, \nu, t_0^{(1)}, t_0^{(2)}, c^*).$$

Data are generated at $\mathcal{R}_0^{(1)} = 2.8$, $\mathcal{R}_0^{(2)} = 3.2$, $\nu = 1.414$, $t_0^{(1)} = 12$, $t_0^{(2)} = 18$, $c^* = 0.3$, and $t_1 = 22$.

This scenario is intended to mimic a more realistic setting in which different regions/countries experience different transmissibility and react at different times, but where the distribution of susceptibility and the long-term strength of NPIs are assumed comparable.

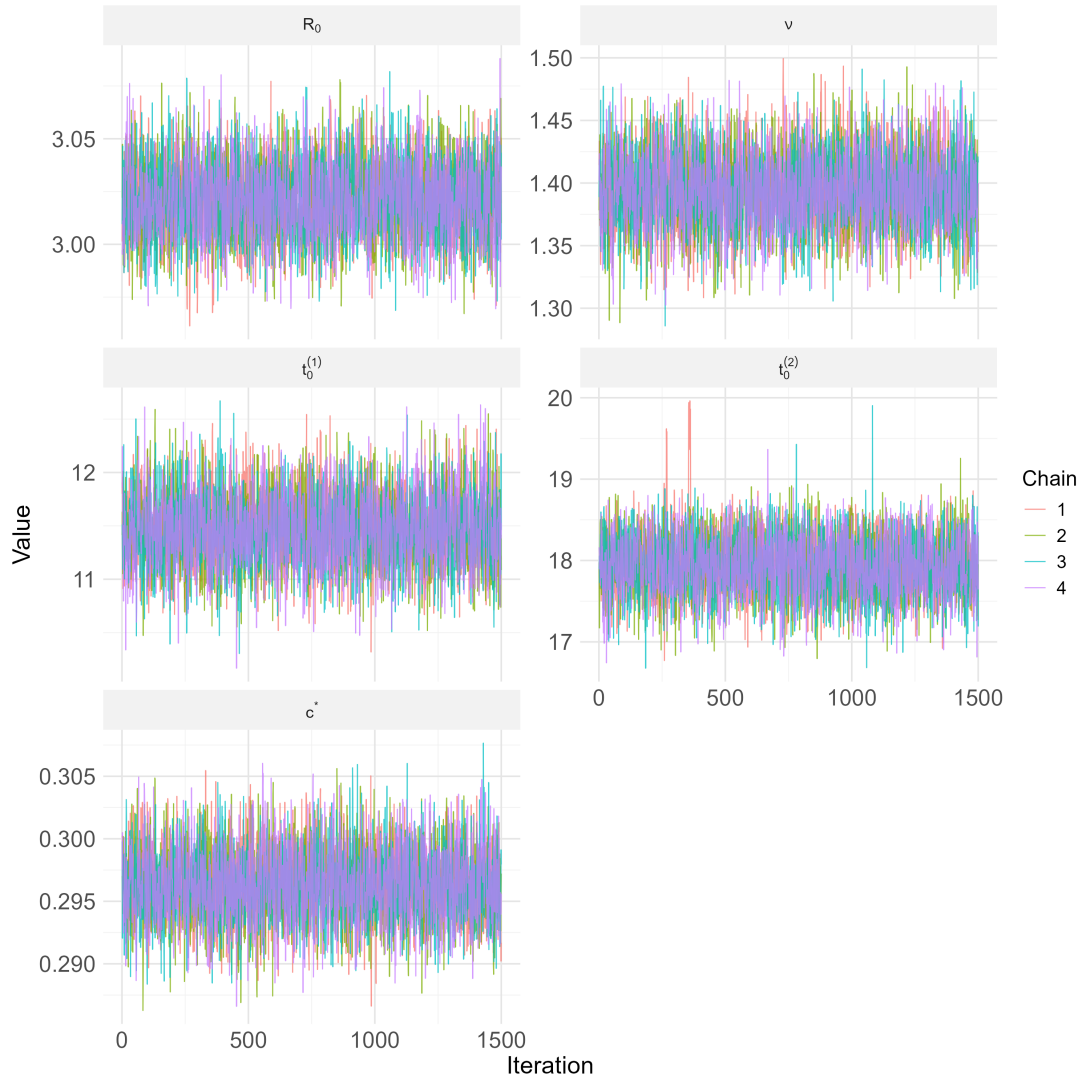


Figure D.8: MCMC trace plots for the 5-parameter two-epidemic model (Scenario A). The shared parameters $\mathcal{R}_0$, ν , and c^* show good mixing and stabilise consistently in the posterior space. The second epidemic-specific onset time $t_0^{(2)}$ centres around its true value, while $t_0^{(1)}$ converges slightly below the true value of 12. The remaining parameters cluster reasonably around their simulation values.

D.4.1 Posterior summaries: Scenario B

Table D.5 reports posterior medians, SDs, and 95% CrIs. Relative to Scenario A (Table D.3, Width for ν is 0.118 and 0.0117 for c^*), posterior uncertainty increases for ν (width 0.208) and c^* (width 0.0210) and the posterior median of ν shifts upward (from 1.394 to 1.496), consistent with a partial return of the ν - c^* compensation ridge when transmission intensity is allowed to vary by epidemic as seen in Figures D.9 where marginal posterior densities are shown. Similarly, The onset times also become less precise: the CrI widths increase from 1.35 and 1.42 days in the five-parameter model to 1.54 and 1.72 days for $t_0^{(1)}$ and $t_0^{(2)}$ respectively. Thus the six-parameter specification still recovers the main structure of the two epidemics, but with noticeably more posterior uncertainty than in Scenario A.

Table D.5: Scenario B (6 parameters): posterior summaries for the two-epidemic model with epidemic-specific reproduction numbers and onset times. Entries are posterior medians with posterior SD and central 95% CrIs. “Width” is the CrI width.

Parameter	True	Median	SD	2.5%	97.5%	Width
$\mathcal{R}_0^{(1)}$	2.800	2.837	0.030	2.778	2.899	0.122
$\mathcal{R}_0^{(2)}$	3.200	3.193	0.021	3.146	3.234	0.088
ν	1.414	1.496	0.053	1.389	1.597	0.208
$t_0^{(1)}$	12.00	12.47	0.39	11.70	13.23	1.54
$t_0^{(2)}$	18.00	18.24	0.40	17.61	19.33	1.72
c^*	0.3000	0.3063	0.0054	0.2961	0.3171	0.0210

D.4.2 Fit quality and diagnostics: Scenario B

The fitted trajectories align closely with the observed case counts (Figure D.10) , and the residual plots shows a stable dispersion around zero (Figure D.11). Furthermore, despite the increased dimensionality of the parameter space, the trace plots demonstrate consistent MCMC convergence and adequate mixing across all four chains (Figure D.12).

D.4.3 Re-emergence of the ν – c^* ridge: Scenario B

The main qualitative difference between Scenarios A and B lies in the correlation structure of the posterior. The selected correlations for the six-parameter fit are summarised in Table D.6. Two dependencies are most relevant to the identifiability question.

First, there is a trade-off between $R_0^{(2)}$ and $t_0^{(2)}$, with $\text{corr}(R_0^{(2)}, t_0^{(2)}) \approx -0.83$. For the later epidemic, a higher reproduction number can be partially balanced by an earlier onset of behavioural change, and vice versa, without strongly altering the observed trajectory. A weaker negative association of the same type is visible between $R_0^{(1)}$ and $t_0^{(1)}$.

Second, and more important for the identifiability question, the correlation between the heterogeneity parameter and the NPI floor becomes very strong: $\text{corr}(\nu, c^*) \approx 0.92$. This is higher than in Scenario A, where $\text{corr}(\nu, c^*) \approx 0.71$ (Table D.4), and much closer to the strong compensation seen in the single-epidemic analyses of Chapter 2, $\text{corr}(\nu, c^*) \approx 0.97$. Allowing both R_0 and t_0 to vary by epidemic effectively “uses up” much of the cross-epidemic contrast that previously helped to pin down (ν, c^*) , so that changes in heterogeneity and in long-term NPI strength can once again substitute for one another in the sense of trajectories.

The strong ν – c^* correlation (0.92) indicates that the multi-epidemic approach may compromise identifiability gains when epidemic-specific flexibility is allowed. With both transmission rates and onset times varying by epidemic, the shared parameters ν and c^* have less constraining

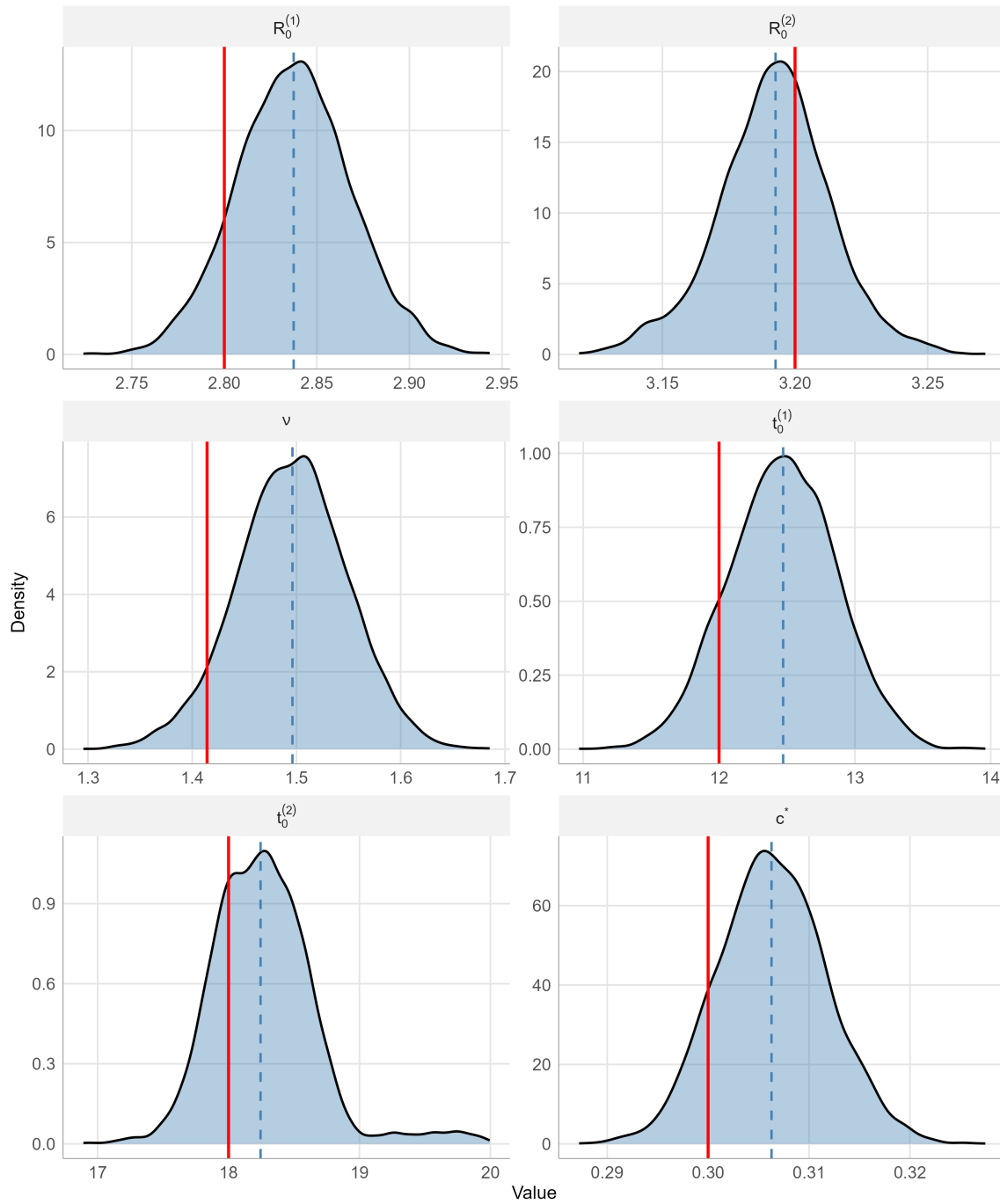


Figure D.9: Marginal posterior densities for the 6-parameter model (Scenario B). Solid red vertical lines mark the true data-generating values, while dashed blue lines indicate the posterior medians. While the model captures the true parameters, the broader spread of the distributions for ν and c^* reflects increased posterior uncertainty.

information from cross-epidemic comparison. The model can fit observed trajectories by trading off heterogeneity against NPI strength nearly as freely as in single-epidemic analysis.

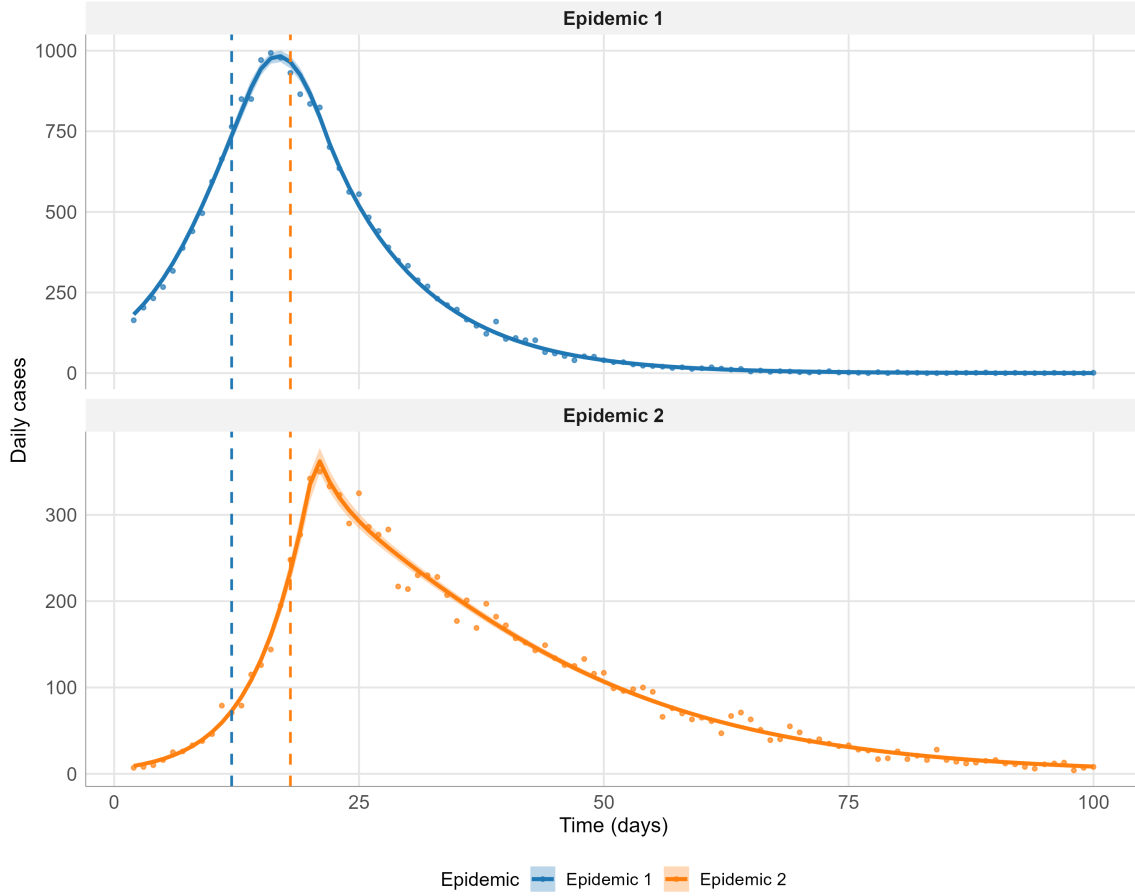


Figure D.10: Fitted versus observed daily incidence under the 6-parameter model (Scenario B). Scatter points represent the observed daily cases for Epidemic 1 (top) and Epidemic 2 (bottom). Solid lines trace the fitted posterior medians, surrounded by 95% posterior predictive ribbons. The vertical dashed lines mark the true, epidemic-specific onset times of behavioural change ($t_0^{(1)}$ and $t_0^{(2)}$).

Table D.6: Selected posterior correlations for the 6-parameter model (Scenario B). The table reports the main pairs discussed in the text: the $\mathcal{R}_0$ – t_0 timing trade-off and the re-emergent ν – c^* compensation ridge.

Parameter pair	Posterior correlation
$\mathcal{R}_0^{(1)}-t_0^{(1)}$	-0.15
$\mathcal{R}_0^{(2)}-t_0^{(2)}$	-0.83
ν – c^*	0.92

D.5 Appendix conclusion

The result in section D.2 supports the main Chapter 3 conclusion: the two-epidemic model improves practical identifiability of the shared parameters, especially along the ν – c^* compensation direction. In this representative run, the 95% CrI widths shrink by factors of about 3.93 ($\mathcal{R}_0$), 6.89 (ν), 2.20 (t_0), and 8.86 (c^*), and the ν – c^* posterior correlation drops markedly when moving from one to two epidemics (Table D.2). These results are comparable to the profile likelihood and

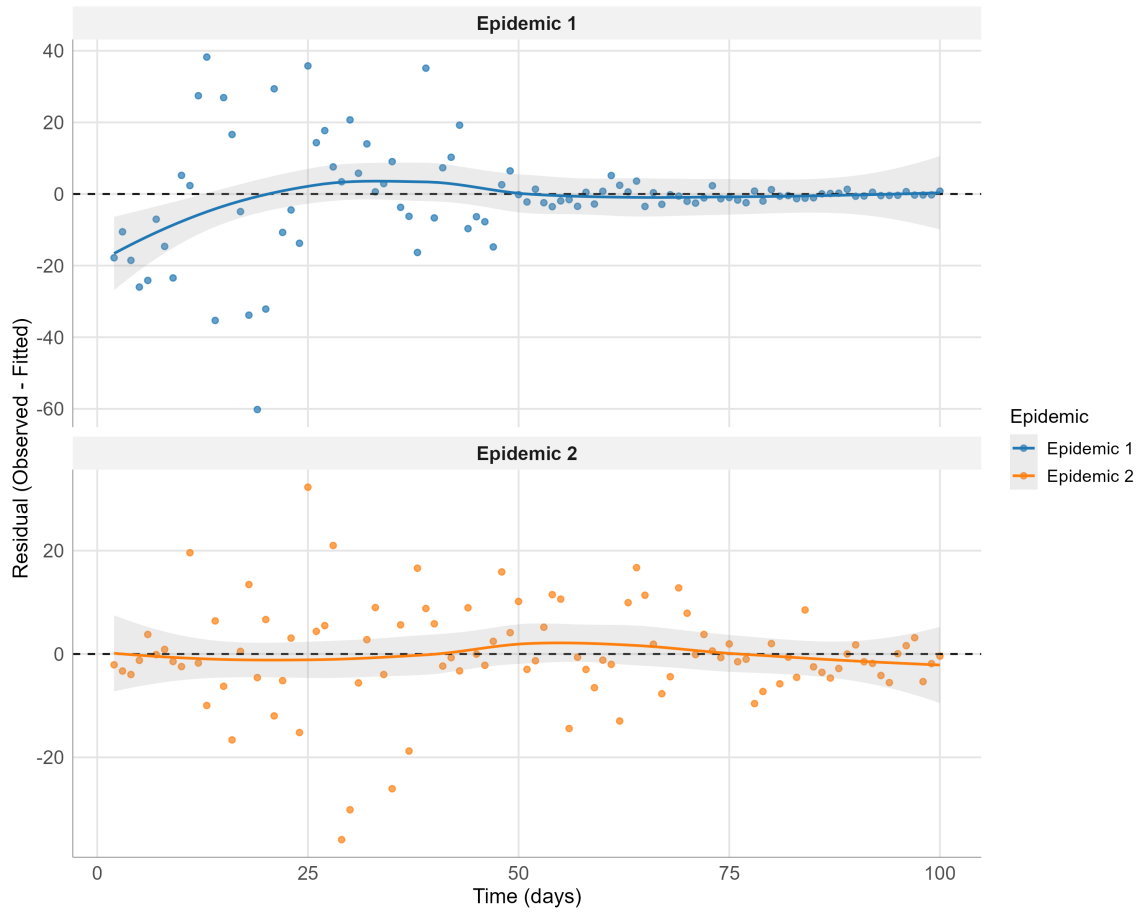


Figure D.11: Residuals versus time for the 6-parameter model, calculated as observed incidence minus the fitted posterior median. The solid lines represent smoothed residual trends with 95% uncertainty bands. RMSE values are 14.11 for Epidemic 1 and 9.71 for Epidemic 2.

Hessian-based results in the main text. Scenarios A and B provide a controlled demonstration of the realism–identifiability trade-off. Allowing epidemic-specific onset times of behavioural change (Scenario A) retains much of the identifiability benefit of the basic two-epidemic model with fully shared-parameter vector θ , but already increases dependence between ν and c^* . Allowing epidemic-specific transmission intensity as well (Scenario B) reintroduces the ν – c^* compensation ridge and widens the credible intervals for these parameters.

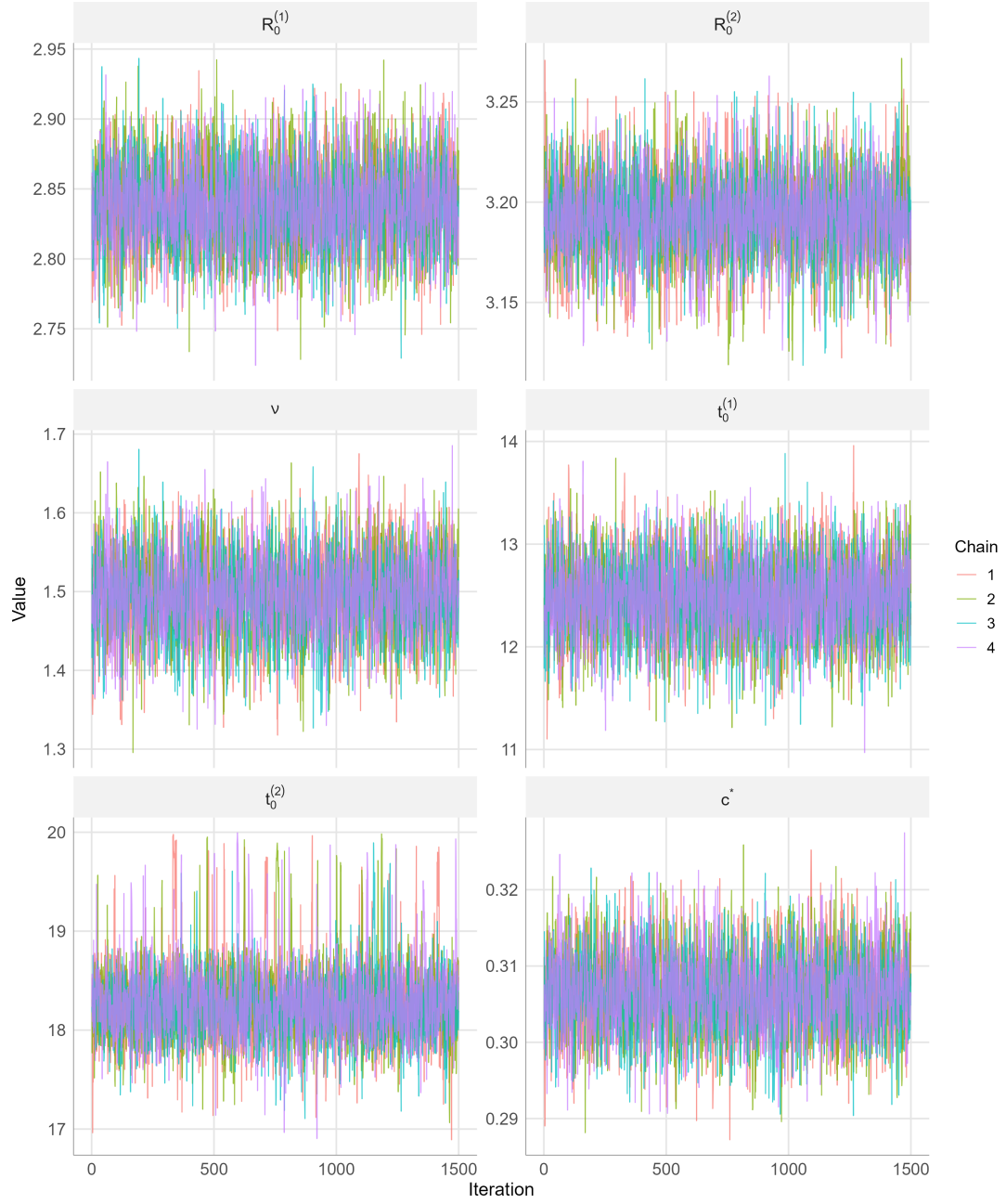


Figure D.12: Trace plots for the 6-parameter model (Scenario B) across four parallel MCMC chains. The sampled trajectories for $\mathcal{R}_0^{(1)}$, $\mathcal{R}_0^{(2)}$, ν , $t_0^{(1)}$, $t_0^{(2)}$, and c^* show good overlap, stable behaviour over iterations, and no visible long-term drift or chain separation, suggesting satisfactory mixing and convergence of the sampler.

Appendix E

Extension of Discretisation

Methodology and Distribution-Level Validation

E.1 Introduction and Motivation

This appendix extends the discretisation framework developed in Chapter 4 from the Gamma family to the Lognormal family. The goal is to construct a mean-one, moment-matched discrete approximation that can be used in the finite susceptibility-stratified SEIR system (4.3), while preserving positivity and numerical stability across a range of heterogeneity levels.

We first parameterise a Lognormal distribution to have unit mean and prescribed coefficient of variation ν , then apply the same three-stage pipeline used for Gamma: CDF binning, conditional-mean representatives, and moment transformation (using LA or SM). Finally, we summarise distribution-level diagnostics that complement the trajectory-level validation in the main chapter and provide the resolution checks needed for the Gamma–Lognormal comparisons in Chapter 5.

E.2 The Lognormal Distribution with unit mean

Let $Y \sim \mathcal{N}(\mu, \sigma^2)$ and $X = \exp(Y)$. Then for any $t \in \mathbb{R}$,

$$\mathbb{E}[e^{tY}] = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \quad (\text{E.1})$$

which is the moment generating function of a normal distribution evaluated at t .

Setting $t = 1$ and $t = 2$ we obtain the first two raw moments of X :

$$\mathbb{E}[X] = \mathbb{E}[e^Y] = \exp\left(\mu + \frac{1}{2}\sigma^2\right), \quad \mathbb{E}[X^2] = \mathbb{E}[e^{2Y}] = \exp\left(2\mu + 2\sigma^2\right). \quad (\text{E.2})$$

Hence,

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \\ &= (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}. \end{aligned} \quad (\text{E.3})$$

The coefficient of variation (CV) follows immediately:

$$\text{CV}(X)^2 = \frac{\text{Var}(X)}{\mathbb{E}[X]^2} = \frac{(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}}{e^{2\mu+\sigma^2}} = e^{\sigma^2} - 1, \quad (\text{E.4})$$

$$\Rightarrow \text{CV}(X) = \sqrt{e^{\sigma^2} - 1}. \quad (\text{E.5})$$

Setting up for mean = 1 and prescribed CV = ν

1. Fix the mean at 1.

Choose

$$\mu = -\frac{1}{2}\sigma^2 \quad \Rightarrow \quad \mathbb{E}[X] = \exp\left(-\frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2\right) = 1.$$

2. Hit a desired CV.

From $\text{CV}^2 = e^{\sigma^2} - 1$ we solve

$$\sigma^2 = \log(1 + \nu^2).$$

Thus the parameter pair

$$\boxed{\sigma^2 = \log(1 + \nu^2), \quad \mu = -\frac{1}{2}\sigma^2} \quad (\text{E.6})$$

ensures that

$$\mathbb{E}[X] = 1, \quad \text{Var}(X) = \nu^2.$$

The discretisation procedure described below approximates such a mean-one lognormal distribution with k discrete classes while preserving its moments exactly. We apply the same three-stage pipeline as in Chapter 4; details of the lognormal discretisation steps are given in Section E.3.

E.2.1 Representative Value Selection

Bin edges. Let $0 = t_0 < t_1 < \dots < t_k = +\infty$ be the quantile edges obtained from the Stage 1 CDF binning step, with $t_i = F_{\mu,\sigma}^{-1}(c_i)$ and $c_i = \sum_{j \leq i} q_j$. Define $a_i = t_{i-1}$ and $b_i = t_i$ so that the i th bin is $[a_i, b_i)$. Within each interval $[a_i, b_i)$, a representative value x_i is chosen to summarize the local behaviour of the distribution. We adopt the conditional expectation within each bin, which minimizes bias in the discrete approximation:

$$x_i = \mathbb{E}[X \mid a_i \leq X < b_i] = \frac{\int_{a_i}^{b_i} x f_{\mu,\sigma}(x) dx}{F_{\mu,\sigma}(b_i) - F_{\mu,\sigma}(a_i)}. \quad (\text{E.7})$$

For the lognormal distribution, $X \sim \text{Lognormal}(\mu, \sigma^2)$ implies $\ln X \sim \mathcal{N}(\mu, \sigma^2)$ with density

$$f_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0.$$

The cumulative distribution function (CDF) and its inverse quantile function are given by

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \quad F_{\mu,\sigma}^{-1}(p) = \exp(\mu + \sigma \Phi^{-1}(p)), \quad (\text{E.8})$$

where $\Phi(\cdot)$ is the standard normal CDF,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt, \quad (\text{E.9})$$

and $\Phi^{-1}(\cdot)$ denotes its inverse quantile function.

Deriving the conditional expectation. The numerator of the conditional mean is the partial first moment over the interval $[a_i, b_i)$:

$$\mathbb{E}[X; a_i \leq X < b_i] = \int_{a_i}^{b_i} x f_{\mu,\sigma}(x) dx.$$

Applying the change of variable $y = \ln x$ (so that $x = e^y$ and $dx = e^y dy$), the limits become $y_a = \ln a_i$ and $y_b = \ln b_i$, yielding

$$\mathbb{E}[X; a_i \leq X < b_i] = \frac{1}{\sigma\sqrt{2\pi}} \int_{y_a}^{y_b} \exp\left(y - \frac{(y - \mu)^2}{2\sigma^2}\right) dy.$$

Completing the square in the exponent gives

$$y - \frac{(y - \mu)^2}{2\sigma^2} = -\frac{(y - (\mu + \sigma^2))^2}{2\sigma^2} + \mu + \frac{1}{2}\sigma^2,$$

so that

$$\mathbb{E}[X; a_i \leq X < b_i] = e^{\mu + \frac{1}{2}\sigma^2} \frac{1}{\sigma\sqrt{2\pi}} \int_{y_a}^{y_b} \exp\left(-\frac{(y - (\mu + \sigma^2))^2}{2\sigma^2}\right) dy.$$

Recognizing the remaining integral as a difference of standard normal CDF values, we obtain

$$\mathbb{E}[X; a_i \leq X < b_i] = e^{\mu + \frac{1}{2}\sigma^2} \left[\Phi\left(\frac{\ln b_i - (\mu + \sigma^2)}{\sigma}\right) - \Phi\left(\frac{\ln a_i - (\mu + \sigma^2)}{\sigma}\right) \right]. \quad (\text{E.10})$$

Closed-form conditional mean. Dividing by the probability mass of the bin,

$$q_i = F_{\mu, \sigma}(b_i) - F_{\mu, \sigma}(a_i) = \Phi\left(\frac{\ln b_i - \mu}{\sigma}\right) - \Phi\left(\frac{\ln a_i - \mu}{\sigma}\right),$$

we obtain the closed-form expression for the conditional mean:

$$x_i = \frac{e^{\mu + \frac{1}{2}\sigma^2} \left[\Phi\left(\frac{\ln b_i - (\mu + \sigma^2)}{\sigma}\right) - \Phi\left(\frac{\ln a_i - (\mu + \sigma^2)}{\sigma}\right) \right]}{\Phi\left(\frac{\ln b_i - \mu}{\sigma}\right) - \Phi\left(\frac{\ln a_i - \mu}{\sigma}\right)}. \quad (\text{E.11})$$

The selection of x_i as the true conditional mean ensures that the discretised grid approximates the underlying lognormal mean structure accurately before any subsequent correction or transformation.

E.3 Lognormal discretisation: distribution-specific ingredients

We instantiate the distribution-agnostic discretisation pipeline of Section 4.2 for a mean-one lognormal target with prescribed coefficient of variation (CV) ν . As for Gamma, the pipeline itself is unchanged; here we record only the lognormal-specific ingredients (parameterisation, quantiles, and bin representatives), and then apply the generic transformation step (LA or SM).

Parameterisation. We use the mean-one, prescribed-CV parameterisation derived in Appendix E.2:

$$\sigma^2 = \log(1 + \nu^2), \quad \mu = -\frac{1}{2}\sigma^2,$$

so that $X \sim \text{Lognormal}(\mu, \sigma^2)$ satisfies $\mathbb{E}[X] = 1$ and $\text{Var}(X) = \nu^2$.

CDF binning (lognormal quantiles). Stage 1 is identical to the Gamma case: choose bin probabilities $\{q_i\}$ using equal or linear CDF binning, form cumulative masses $c_i = \sum_{j \leq i} q_j$, and

set lognormal quantile edges

$$t_i = F_{\mu,\sigma}^{-1}(c_i) = \exp(\mu + \sigma \Phi^{-1}(c_i)), \quad a_i = t_{i-1}, \quad b_i = t_i,$$

with $t_0 = 0$ and $t_k = +\infty$.

Representatives (bin conditional means). Stage 2 uses within-bin conditional expectations, with the closed-form expression (E.11) for each bin $[a_i, b_i)$. We then mean-normalise the initial grid:

$$x \leftarrow \frac{x}{\sum_i q_i x_i}.$$

Moment transformation (LA or SM). Finally, we apply the distribution-agnostic transformation step of Section 4.2 with $(\mu_\star, \nu^2) = (1, \nu^2)$. In the implementation used in this thesis: (i) LA returns the mean-normalised initial discretisation if the root cannot be bracketed; and (ii) SM returns the mean-normalised initial discretisation (identity multipliers) when no admissible solution exists.

Numerical illustrations for LA and SM in the lognormal case are provided in Section E.4.

E.4 Numerical illustration for the Lognormal discretisation

We illustrate the lognormal discretisation with $k = 10$ and $\text{CV} = 1.155$. Parameters are set to ensure mean one: $\sigma^2 = \log(1 + \text{CV}^2)$ and $\mu = -\frac{1}{2}\sigma^2$. We consider two CDF binnings (equal and linear); bin representatives are conditional expectations. Transformation uses the two methods from Section 4.2: log-affine (LA) and split-multiplicative (SM).

LA: scalar root and shape preservation

The LA map solves

$$g(B) = h(2B) - 2h(B) - \log(1 + \nu^2) = 0, \quad h(t) = \log\left(\sum_i q_i x_i^t\right),$$

yielding a unique $B^\star$. With linear binning we obtain $B^\star \approx 1.0736$, and the corresponding shift $A^\star$ enforces the mean. The monotone $g(B)$ curve and its root are shown in Fig. E.1; the rectangle-PDF overlay after LA transformation appears in Fig. E.2. Across both binnings, LA keeps the grid ordered and close to the original representatives while matching the target moments.

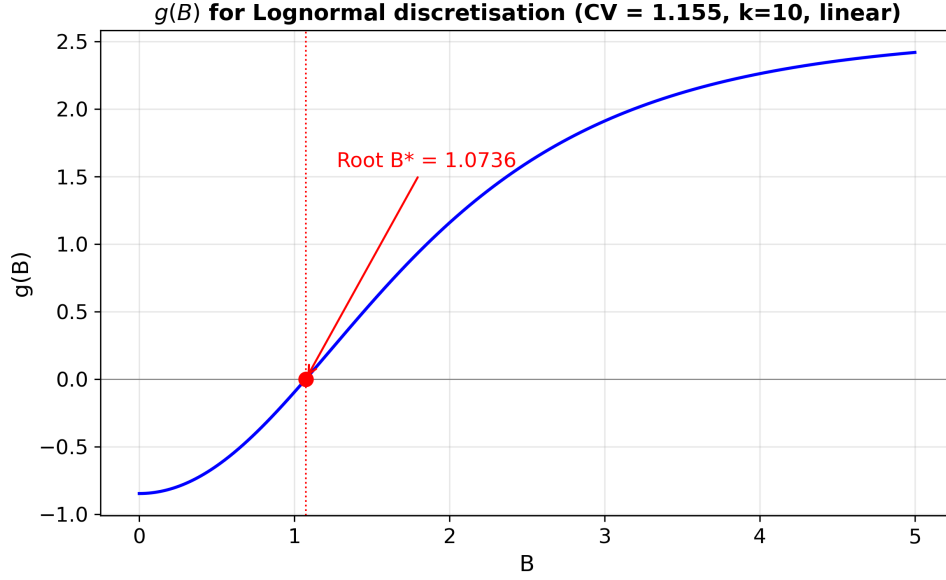


Figure E.1: $g(B)$ for the discretised log-normal ($CV = 1.155$). The curve increases monotonically and crosses zero at $B^* \approx 1.0736$ (linear spacing).

Table E.1: Log-affine transformation parameters and empirical checks for the log-normal case ($CV = 1.155$, $k = 10$).

Spacing	A^*	B^*	Emp. Mean	Emp. Var
Linear	-0.032	1.074	1.000	1.334
Equal	-0.076	1.173	1.000	1.334

SM: two-factor rescaling with admissible-root selection

SM splits the grid at the median mass and applies positive multipliers (r, u) to the two halves, chosen from the quadratic's admissible roots, followed by a tiny global renormalisation. For this CV the selected multipliers are:

Table E.2: Split-multiplicative transformation parameters and empirical checks for the log-normal case ($CV = 1.155$, $k = 10$).

Spacing	r	u	Emp. Mean	Emp. CV
Linear	0.726	1.060	1.000	1.155
Equal	0.494	1.110	1.000	1.155

Visual comparison of calibrated representatives

Figures E.3 and E.4 compare the original conditional-mean representatives with the calibrated grids produced by the LA and SM transformations under equal and linear spacing. In both cases, the calibrated representatives stay close to the original grid and preserve the increasing order

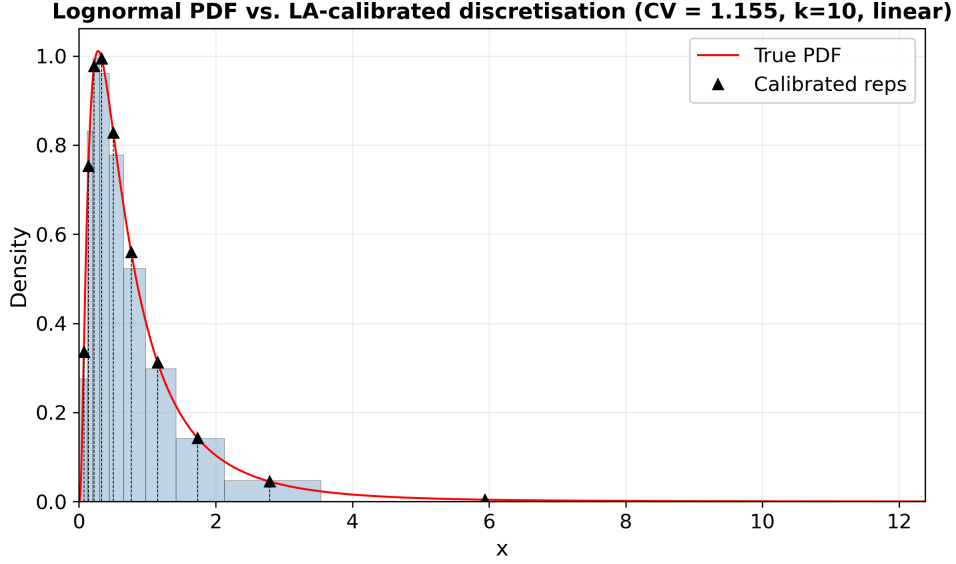


Figure E.2: Log-normal PDF vs. discretised rectangles after LA transformation ($CV = 1.155$, $k = 10$, linear spacing). Rectangles depict bin mass/width; dashed stems indicate the calibrated x'_i .

across groups. This shows that moment matching is achieved without materially changing the structure of the discrete lognormal approximation.

The equal-spacing grid shows slightly larger corrections in the upper tail, which is expected because the lognormal distribution allocates more mass to large values in the final bins. The linear-spacing grid requires less visible adjustment, with the calibrated and original curves almost overlapping throughout. In both settings, LA and SM behave similarly, although LA gives a smoother adjustment across all groups.

E.5 Distribution-Level Validation

E.5.1 Overview

This appendix validates the discretisation framework of Section 4.2 at the distribution level: how well does the discrete representation approximate the shape of its continuous target? Since the LA and SM transformation stages (Sections 4.2.2 and 4.2.1) enforce the target mean ($E[X] = 1$) and target variance to machine precision whenever an admissible solution exists, the mean and CV errors are effectively zero by construction. The remaining diagnostic question is therefore how closely the *probability mass allocation* of the discrete grid matches the continuous density, which we quantify through the density error defined below.

We evaluated all combinations of the following factors:

- **Distributions:** Gamma and Lognormal.

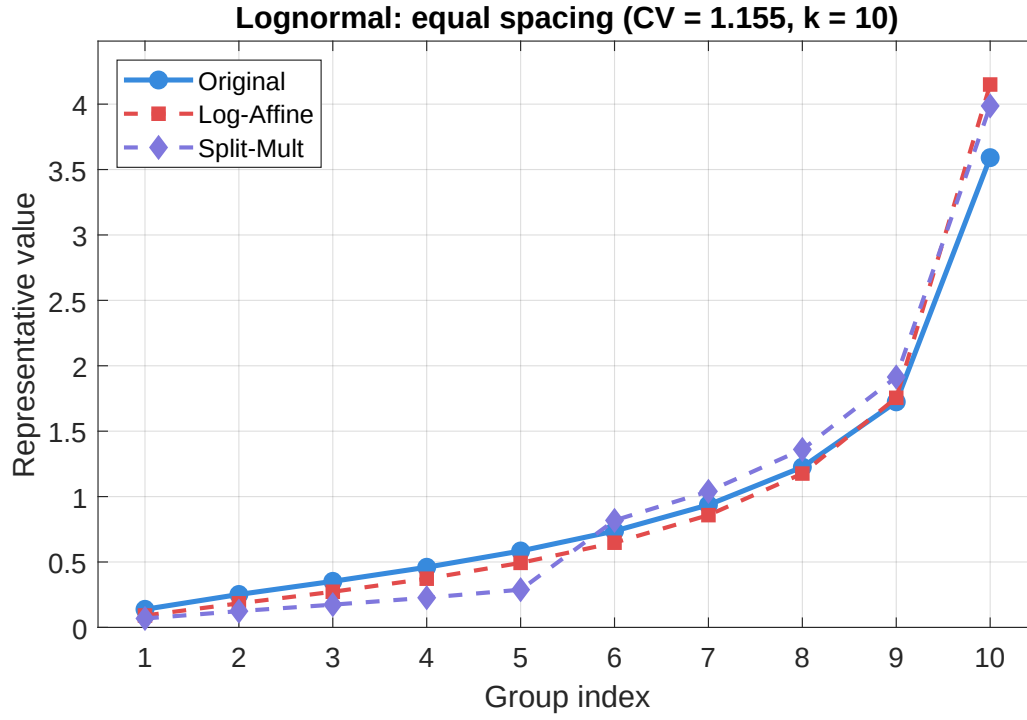


Figure E.3: Lognormal discretisation with equal probability spacing ($CV = 1.155$, $k = 10$). The blue solid curve shows the original conditional-mean representatives, while the red dashed and purple dashed curves show the calibrated representatives obtained using the log-affine (LA) and split-multiplicative (SM) transformations, respectively. Both transformations preserve positivity and ordering, with the main adjustments occurring in the upper susceptibility groups where matching the target variance requires stronger correction.

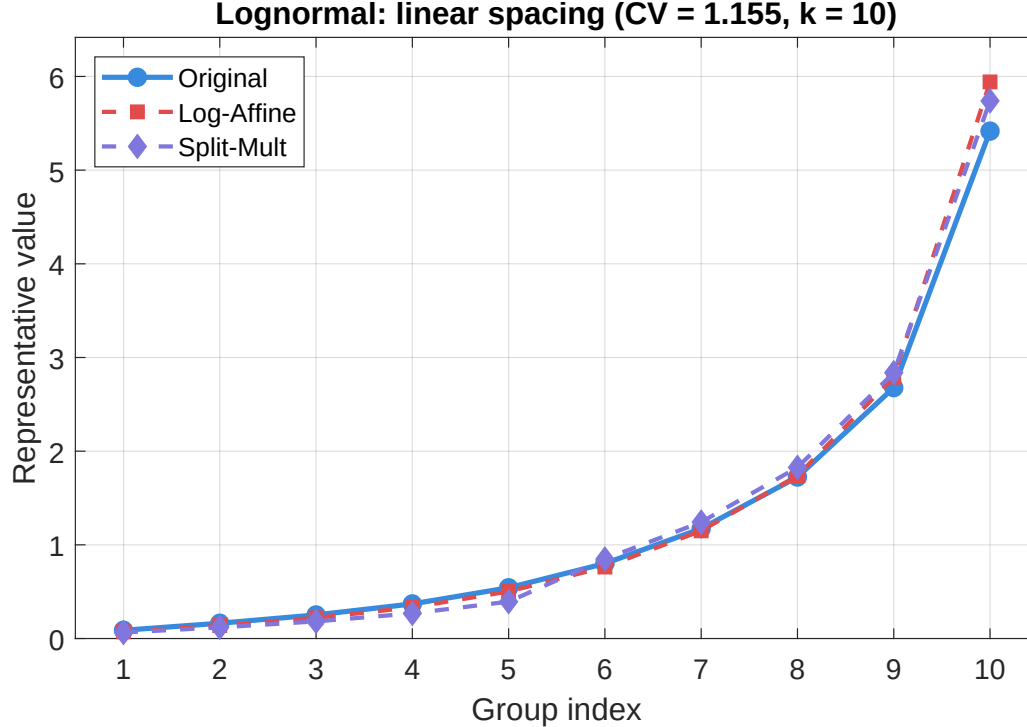


Figure E.4: Lognormal discretisation with linear probability spacing ($CV = 1.155$, $k = 10$). The blue solid curve shows the original conditional-mean representatives, while the red dashed and purple dashed curves show the calibrated representatives from the log-affine (LA) and split-multiplicative (SM) transformations, respectively. The three curves are nearly indistinguishable over most groups, showing that both calibration methods preserve the shape of the initial discretisation while enforcing the target mean and coefficient of variation.

- **Transformation:** Log-Affine (LA, Section 4.2.2) and Split-Multiplicative (SM, Section 4.2.1).
- **CDF binning:** Linear and Equal spacing.
- **Heterogeneity:** $CV \in \{0.5, 1.0, 1.5, 2.0\}$.
- **Resolution:** $k \in \{10, 20, 30, 40, 50, 100\}$.

E.5.2 Density Error Metric

Let the discretisation produce k groups with probability weights $q_1, \dots, q_k$ and calibrated representatives $x_1, \dots, x_k$. We construct bin boundaries $0 = b_0 < b_1 < \dots < b_k$ via the midpoints

$$b_i = \frac{x_i + x_{i+1}}{2}, \quad i = 1, \dots, k-1, \quad (\text{E.12})$$

with $b_0 = 0$ and the upper boundary

$$b_k = \max(2x_k, F^{-1}(0.999), F^{-1}(F(b_{k-1}) + q_k)),$$

where F denotes the target CDF. This ensures the final bin captures essentially all remaining probability mass. Writing $w_i = b_i - b_{i-1}$ for the bin widths, the empirical and theoretical average densities in bin i are

$$\hat{f}_i = \frac{q_i}{w_i}, \quad \bar{f}_i = \frac{F(b_i) - F(b_{i-1})}{w_i}.$$

The density error is the width-weighted mean-squared relative deviation:

$$D_{\text{err}} = \frac{1}{k} \sum_{i=1}^k \frac{w_i (\hat{f}_i - \bar{f}_i)^2}{\bar{f}_i + \epsilon}, \quad (\text{E.13})$$

where $\epsilon = 10^{-10}$ prevents division by zero. Weighting by w_i ensures that wider bins, which contribute more to the overall distributional shape, have proportionally greater influence.

E.5.3 Numerical Stability at High Variance

Both algorithms are stable across the full parameter grid tested. All $4 \times 6 \times 2 = 48$ combinations per scenario completed successfully, yielding $n = 24$ valid runs per spacing method in every case.

For the SM transformation, identity-multiplier fallbacks ($r=u=1$) occur at high CV when the SM quadratic (Section 4.2.1) yields no admissible positive root within the distortion cap ($|\log \mu| \leq 3$). These runs produce valid density errors but reflect the uncalibrated initial

discretisation rather than a genuinely SM-calibrated solution, which accounts for the moderately elevated SM aggregates in Table E.3.

E.5.4 Aggregate Summary of Density Errors

Table E.3 reports the mean and median of D_{err} (E.13) across all valid runs. For each scenario the aggregate pools density errors from every (k, CV) combination with $k \in \{10, 20, 30, 40, 50, 100\}$ and $\text{CV} \in \{0.5, 1, 1.5, 2\}$, grouped by spacing method ($n = 24$ runs per group in all cases).

Linear spacing consistently outperforms equal spacing by a substantial margin across all scenarios. For the Lognormal distribution, equal spacing struggles to capture the heavy right tail, producing mean density errors an order of magnitude higher than linear spacing under LA transformation. LA transformation with linear spacing achieves the lowest errors overall; SM is comparable at low-to-moderate CV but produces larger errors at high CV due to the identity-multiplier fallback. For this reason, linear spacing with LA calibration is the preferred default when the aim is to preserve distributional shape accurately, especially for heavy-tailed Lognormal targets.

Table E.3: Aggregate density errors by distribution and transformation method. Each entry pools D_{err} (E.13) over all (k, CV) combinations ($n = 24$ per spacing in every scenario).

Scenario	Spacing	Mean	Median	n
Gamma (LA)	Equal	0.000883	0.000192	24
	Linear	0.000128	0.0000143	24
Gamma (SM)	Equal	0.00122	0.000458	24
	Linear	0.000751	0.0000244	24
Lognormal (LA)	Equal	0.0122	0.000891	24
	Linear	0.000846	0.0000385	24
Lognormal (SM)	Equal	0.0177	0.00203	24
	Linear	0.0197	0.000553	24

E.5.5 Error Reduction with Increasing Resolution

Figures E.5 and E.6 show the percentage density error reduction relative to the $k=10$ baseline under LA transformation:

$$R(k) = \frac{D_{\text{err}}(k=10) - D_{\text{err}}(k)}{D_{\text{err}}(k=10)} \times 100\%. \quad (\text{E.14})$$

Density errors can vary over orders of magnitude across distributions, CV values, and k . Reporting percentage reduction relative to $k=10$ provides a consistent, comparable measure of convergence. We adopt $k=10$ as the baseline because it represents the minimal viable

resolution for the discretisation to meaningfully approximate the continuous distribution of the heterogeneous model (4.3). Moving from $k=10$ to $k=20$ yields the largest gains (typically $> 85\%$ error reduction); the curves flatten for $k \geq 30$, indicating diminishing returns. This is consistent with the trajectory-level finding in Section 4.5 that $k \approx 20$ is sufficient for within-family Gamma trajectory fidelity. However, the distribution-level diagnostics reported here are also relevant for the inter-family comparisons in Chapter 5, where finer grids are needed to avoid underestimating the differences between Gamma and Lognormal susceptibility distributions.

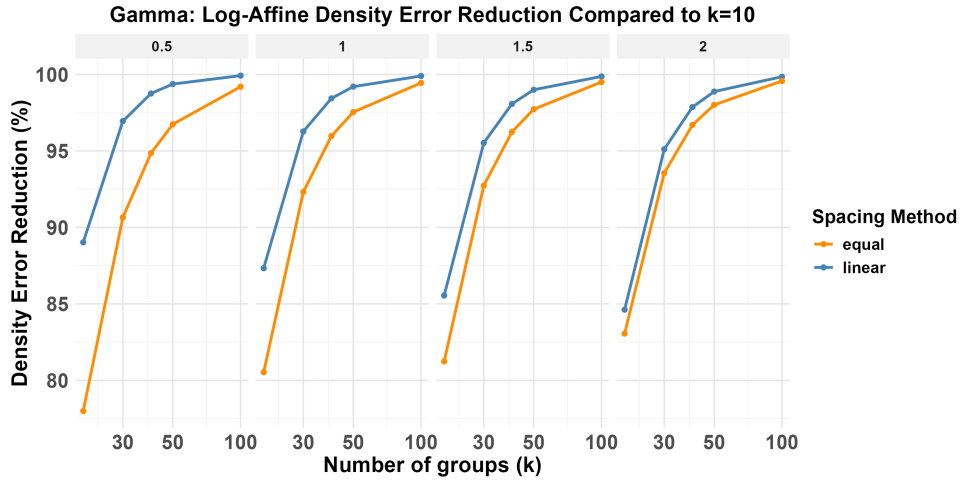


Figure E.5: Percentage reduction in D_{err} for the Gamma distribution (LA transformation) relative to $k=10$. Each panel corresponds to $\text{CV} \in \{0.5, 1, 1.5, 2\}$.

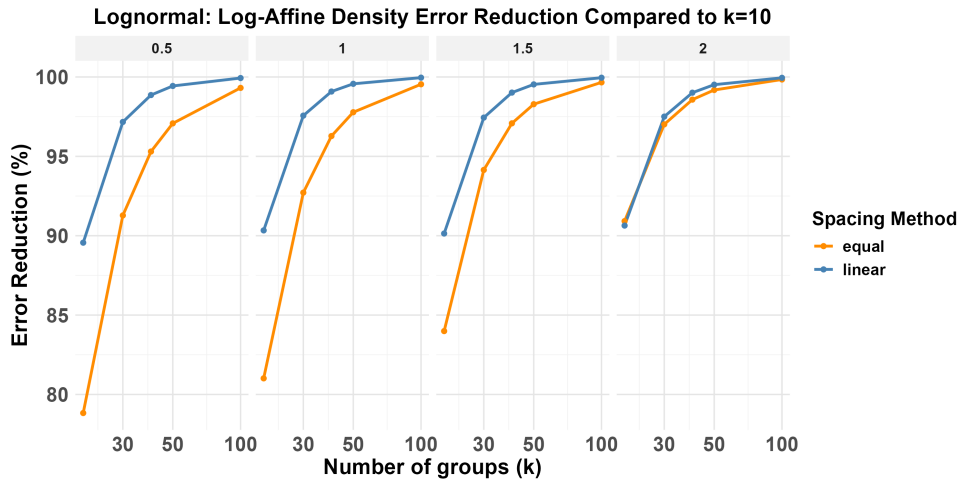


Figure E.6: Percentage reduction in D_{err} for the Lognormal distribution (LA transformation) relative to $k=10$. Linear spacing achieves $> 97\%$ error reduction by $k=50$ even at $\text{CV} = 2$.

Hence, the practical resolution depends on the modelling goal: moderate grids ($k \approx 20$) are adequate for moment matching and within-family trajectory reproduction, whereas inter-family shape comparisons are more reliably assessed on finer grids.

E.6 Summary

This appendix extended the discretisation framework of Chapter 4 from the Gamma family to the Lognormal family. Using the mean-one parameterisation

$$\sigma^2 = \log(1 + \nu^2), \quad \mu = -\frac{1}{2}\sigma^2,$$

we showed that the same three-stage pipeline: CDF binning, conditional-mean representatives, and moment transformation can be applied to Lognormal susceptibility distributions while preserving positivity and matching the target first two moments.

The numerical illustrations in Section E.4 showed that both the log-affine (LA) and split-multiplicative (SM) transformations preserve the ordering of the discrete grid and keep the calibrated representatives close to the original conditional means. For the moderate heterogeneity example considered ($k = 10$, $\text{CV} = 1.155$), both methods performed well, with LA giving a particularly smooth adjustment across the grid.

The distribution-level validation in Section E.5 then examined how well the discretised grids reproduce the shape of their continuous targets. The main conclusion is that spacing choice matters. Linear spacing consistently gave smaller density errors than equal spacing, and this advantage was especially clear for the Lognormal family, where equal spacing struggled to capture the heavier right tail. Thus, LA with linear spacing gave the best performance across the tested configurations, while SM remained usable but was less robust at high heterogeneity because of identity-multiplier fallback in some cases.

Increasing the number of groups improved the distributional approximation, with the largest gain usually occurring when moving from $k = 10$ to $k = 20$. Beyond this, the error curves flattened, indicating diminishing returns. However, this should be interpreted according to the modelling aim. For within-family purposes such as moment matching and trajectory reproduction, a moderate resolution can already be adequate. For inter-family comparisons between Gamma and Lognormal distributions, finer grids are preferable, since Chapter 4 shows that coarse or moderate resolutions can underestimate the magnitude of the difference between the two distributional families.

To sum it up, the results of this appendix support the practical defaults used in the later chapters: conditional-mean representatives, LA transformation, and linear spacing as the preferred discretisation choice. They also provide the distributional validation needed to justify the cross-family robustness and misspecification studies in Chapter 5.

Appendix F

Supplementary results: distribution misspecification inference for the Gamma baseline

This appendix reports the misspecification experiment of Chapter 5 when the data-generating (*baseline*) susceptibility distribution is *Gamma* at high heterogeneity ($\nu = 1.414$). As in the main chapter, each baseline dataset is fitted with three candidate models: a correctly specified Gamma model, a misspecified Lognormal model, and a homogeneous reference model ($\nu = 0$) and inference across replicated datasets is summarised by the median (SD) of the MLEs, the mean width of 95% confidence intervals, and empirical coverage.

F.1 Inference for Gamma-distributed susceptibility

We now discuss the scenario where the baseline population susceptibility follows a Gamma distribution. This is the complement of the Lognormal baseline in the main chapter, and shows that the main conclusions do not depend on which family generates the data.

F.1.1 Single-epidemic inference: Gamma baseline

Table F.1 presents single-epidemic inference for Gamma-generated data at $\nu = 1.414$, with all three fits.

The pattern reflects the Lognormal baseline of the main text. The correctly specified Gamma fit centres near the data-generating value: the median estimate of ν is 1.386, with mean 95% Wald limits (1.061, 1.704) and relative bias -2.0% . Nevertheless, the fit remains weakly informative

for ν and c^* , with wide intervals and sub-nominal coverage.

The misspecified Lognormal fit shifts the median estimate of ν upward to 1.607, with mean 95% Wald limits (1.129, 3.076), relative bias +13.7%, and coverage 65.0%. This is the opposite direction to the Lognormal-baseline case, while the estimates of $\mathcal{R}_0$, t_0 , and c^* remain close to their data-generating values.

The homogeneous fit produces a median estimate of $c^* = 0.239$, with mean 95% Wald limits (0.235, 0.242), relative bias -20.5% , and coverage 0.0%. It also biases $\mathcal{R}_0$ downward, with coverage 20.0%, distinguishing it qualitatively from both heterogeneous fits regardless of which family is correctly specified (Figure F.1).

Table F.1: Single-epidemic inference for Gamma-generated data ($\nu = 1.414$). Median (SD), mean 95% CI width, and empirical coverage across 200 replicates. Convergence (positive-definite Hessian): Gamma 189/200, Lognormal 177/200, homogeneous 200/200. The homogeneous fit fixes $\nu = 0$.

Model	Parameter	Median (SD)	CI width	Coverage (%)
Gamma (correct)	$\mathcal{R}_0$	2.996 (0.024)	0.092	94.2
	ν	1.386 (0.225)	0.643	88.4
	t_0	14.995 (0.337)	1.259	95.2
	c^*	0.296 (0.019)	0.058	86.2
Lognormal (misspecified)	$\mathcal{R}_0$	3.011 (0.033)	0.121	89.3
	ν	1.607 (0.367)	1.946	65.0
	t_0	15.196 (0.385)	1.359	87.6
	c^*	0.304 (0.025)	0.072	78.5
Homogeneous (reference)	$\mathcal{R}_0$	2.943 (0.021)	0.072	20.0
	ν	—	—	—
	t_0	14.912 (0.324)	1.142	95.0
	c^*	0.239 (0.002)	0.008	0.0

F.1.2 Fit and forecast for the Gamma baseline (single epidemic)

Figure F.2 shows how the epidemic trajectory responds to the complete removal of NPIs at day 100 when the baseline distribution is Gamma. The hierarchy seen in the Lognormal baseline fit, Chapter 5 is reproduced: the correctly specified Gamma fit tracks the data-generating trajectory almost exactly; the misspecified Lognormal fit remains close to it but slightly overshoots the second-wave peak; and the homogeneous fit substantially overshoots both the peak height and the final size. Even when the heterogeneous family is misspecified in the heavier-tailed direction, with a Lognormal model fitted to Gamma-generated data, the heterogeneous forecast remains qualitatively closer to the data-generating trajectory than the homogeneous reference.

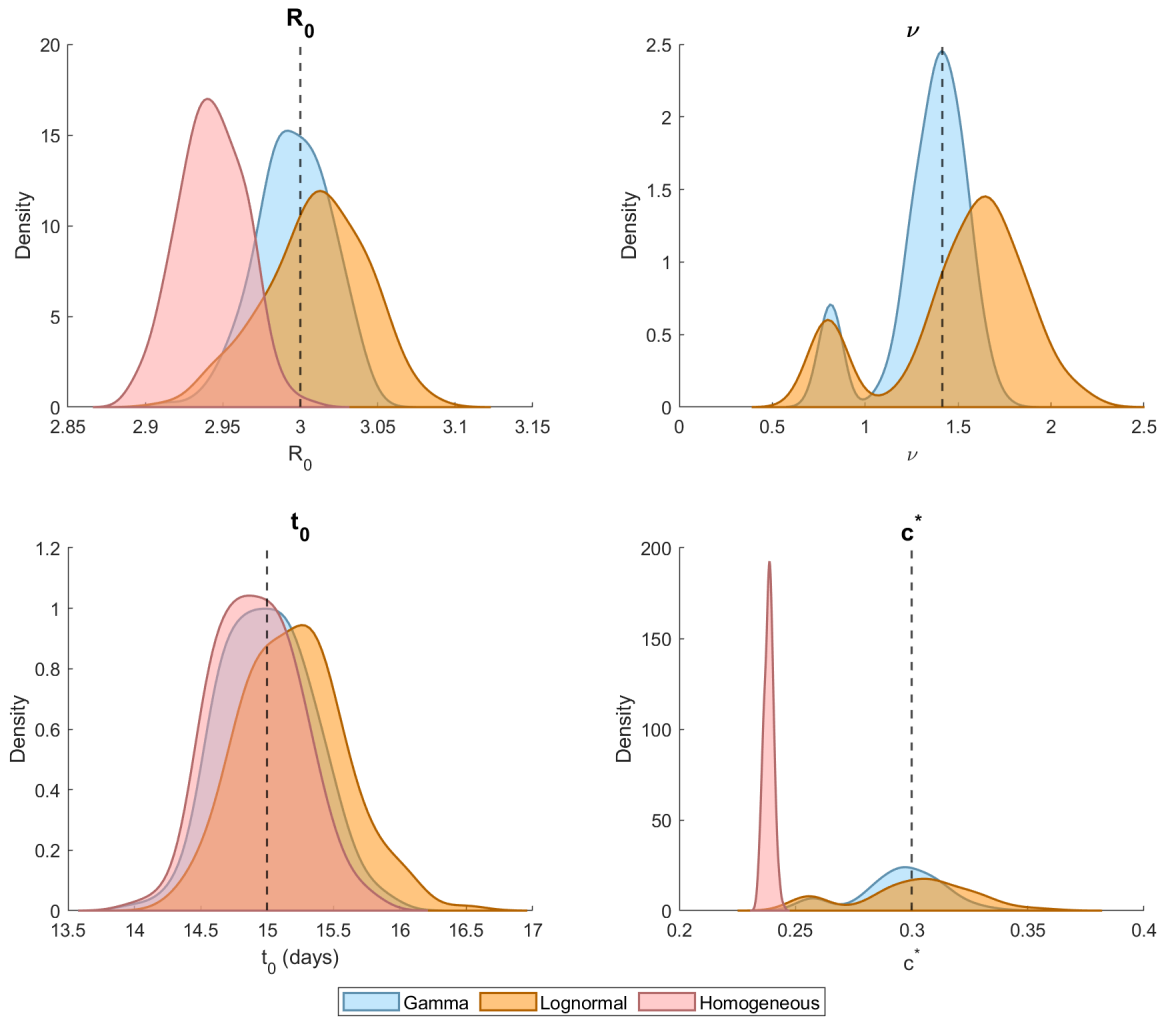


Figure F.1: Distributions of single-epidemic MLE estimates for Gamma-generated data ($\nu = 1.414$). Blue: Gamma (correct). Orange: Lognormal (misspecified). Pink: homogeneous (reference). Dashed vertical lines indicate the data-generating values. The Lognormal density shifts rightward of the data-generating ν but overlaps the Gamma density; the homogeneous fit concentrates around a median c^* estimate of 0.239, with mean 95% Wald limits (0.235, 0.242), well below the data-generating value 0.30.

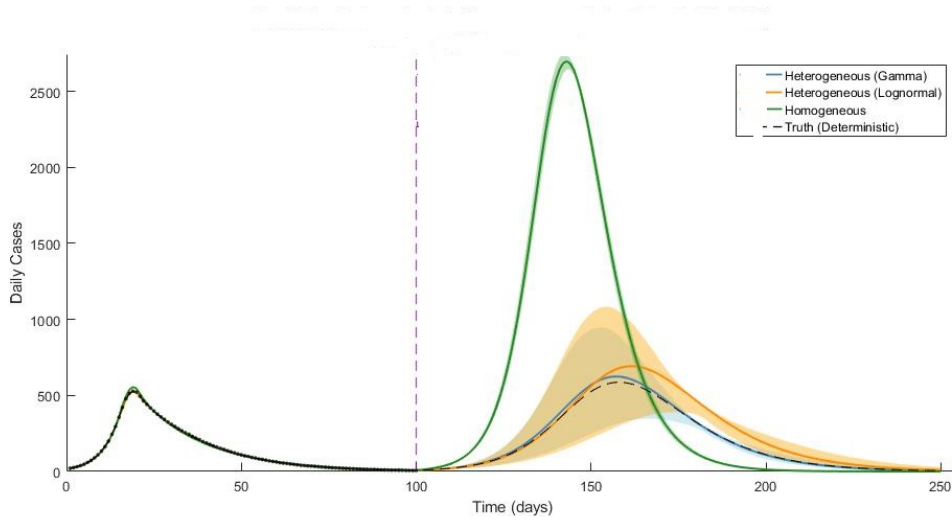


Figure F.2: Single-epidemic forecast trajectories for Gamma-generated data ($\nu = 1.414$, NPIs lifted at day 100). The black dashed line is the deterministic data-generating trajectory; shaded bands are 95% confidence envelopes across replicated fits. The correctly specified Gamma fit (blue) tracks the data-generating trajectory closely. The misspecified Lognormal fit (orange) modestly overshoots the second-wave peak, whereas the homogeneous fit (green) overshoots it substantially.

F.1.3 Two-epidemic inference: Gamma baseline

We now repeat the analysis for the two-epidemic framework. All 200 datasets converged for both heterogeneous fits, giving a cleaner comparison than the single-epidemic case.

Table F.2 shows the identifiability gain for the correct specification. Adding a second epidemic shrinks the CI width for ν from 0.643 to 0.018 and for c^* from 0.058 to 0.005, while coverage for ν rises from 88.4% to 96.5%. The SD for ν drops from 0.225 to 0.005 and the median moves from 1.386 to 1.414, essentially recovering the data-generating value. This mirrors the gain for the Lognormal baseline (Table 5.3) and confirms that joint fitting resolves the practical non-identifiability of ν and c^* regardless of which family generated the data.

Table F.2: Single vs. two-epidemic inference for the correct specification (Gamma-generated data, Gamma fit). Median (SD), mean 95% CI width, and empirical coverage across 200 replicates. Joint fitting sharpens inference for ν and c^* substantially, with coverage rising to near-nominal for all parameters.

Parameter	Single epidemic		Two epidemics		Coverage (%)	
	Median (SD)	Width	Median (SD)	Width	Single	Two
$\mathcal{R}_0$	2.996 (0.024)	0.092	3.001 (0.010)	0.038	94.2	94.0
ν	1.386 (0.225)	0.643	1.414 (0.005)	0.018	88.4	96.5
t_0	14.995 (0.337)	1.259	14.988 (0.112)	0.423	95.2	92.5
c^*	0.296 (0.019)	0.058	0.300 (0.001)	0.005	86.2	90.5

Table F.3 reports all three fits in the two-epidemic setting. When the misspecified Lognormal model is compelled to explain two Gamma-generated epidemics, the compensation ridge that let

it mimic Gamma dynamics in the single-epidemic case can no longer operate. The fit therefore concentrates on biased values.

For the misspecified Lognormal fit, the median estimate of ν is 1.781, with mean 95% Wald limits (1.768, 1.794), relative bias +25.9%, and coverage 0.0%. The median estimate of $\mathcal{R}_0$ is 3.221, with mean 95% Wald limits (3.201, 3.241), relative bias +7.4%, and coverage 0.0%. The median estimate of c^* is 0.321, with mean 95% Wald limits (0.318, 0.323), relative bias +6.8%, and coverage 0.0%.

The homogeneous fit fails differently but equally severely. Its median estimate of $\mathcal{R}_0$ is 2.304, with mean 95% Wald limits (2.296, 2.312); its median estimate of t_0 is 24.882 days, with mean 95% Wald limits (24.785, 24.975), compared with the data-generating value 15; and its median estimate of c^* is 0.291, with mean 95% Wald limits (0.289, 0.294). Coverage is 0.0% for $\mathcal{R}_0$ and t_0 , and 0.5% for c^* . Figure F.3 shows the corresponding MLE densities.

A notable contrast is the behaviour of c^* for the misspecified Lognormal fit: it stays reasonably estimated (0.321, +6.8%) relative to the severe bias in ν , confirming that the structural mismatch is absorbed specifically into the heterogeneity estimate rather than spread across all parameters, as in the single-epidemic case and using the Lognormal baseline.

Table F.3: Two-epidemic inference for Gamma-generated data, fitted with Gamma, Lognormal and homogeneous models ($\nu = 1.414$). Median (SD), mean 95% CI width, and empirical coverage across 200 replicates. Convergence: 200/200 (heterogeneous fits), 187/200 (homogeneous). Under misspecification the Lognormal fit concentrates on biased values for $\mathcal{R}_0$ and ν (0.0% coverage); the direction of the ν bias (+25.9%) is opposite to the Lognormal-baseline case (−15.5%).

Model	Parameter	Median (SD)	CI width	Coverage (%)
Gamma (correct)	$\mathcal{R}_0$	3.001 (0.010)	0.038	94.0
	ν	1.414 (0.005)	0.018	96.5
	t_0	14.988 (0.112)	0.423	92.5
	c^*	0.300 (0.001)	0.005	90.5
Lognormal (misspecified)	$\mathcal{R}_0$	3.221 (0.013)	0.040	0.0
	ν	1.781 (0.009)	0.026	0.0
	t_0	11.812 (0.140)	0.365	0.0
	c^*	0.321 (0.001)	0.004	0.0
Homogeneous (reference)	$\mathcal{R}_0$	2.304 (0.007)	0.016	0.0
	ν	—	—	—
	t_0	24.882 (0.069)	0.189	0.0
	c^*	0.291 (0.002)	0.005	0.5

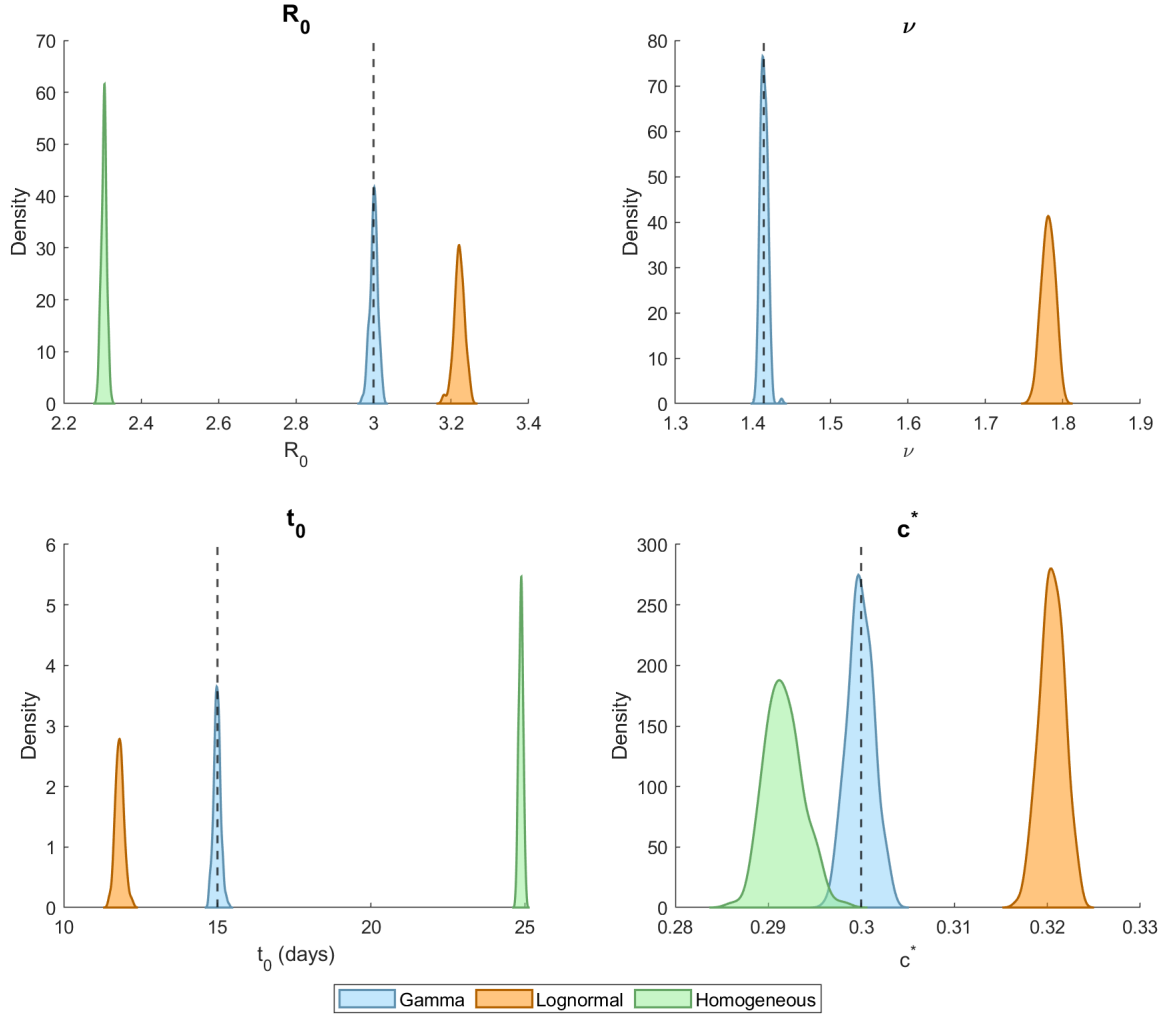


Figure F.3: Distributions of two-epidemic MLE estimates for Gamma-generated data. Blue: Gamma (correct). Orange: Lognormal (misspecified). Green: homogeneous (reference). Dashed vertical lines indicate the data-generating values. The misspecified Lognormal fit collapses to a spike at a median ν estimate of 1.781, with mean 95% Wald limits (1.768, 1.794), well separated from the correctly specified Gamma density. The homogeneous fit fails on R_0 , t_0 and c^* . The direction of the ν shift is opposite to the Lognormal-baseline case (Figure 5.8): a heavier-tailed family fitted to lighter-tailed data inflates the heterogeneity estimate, while a lighter-tailed family fitted to heavier-tailed data deflates it.

F.1.4 Fit and forecast for joint inference (Gamma baseline)

Figure F.4 shows epidemic-trajectory forecasts using the two-epidemic framework, with NPIs lifted from day 100 in each epidemic and parameters estimated jointly from both waves before forecasting. In the training window, all three fits track the observed trajectories closely, and their confidence bands overlap. The forecast period reveals the consequence of the biased heterogeneity estimate. The correctly specified Gamma fit projects the second wave accurately, with a narrow band tracking the data-generating trajectory. The misspecified Lognormal fit, forced by the joint constraint to a median ν estimate of 1.781, with mean 95% Wald limits (1.768, 1.794), assumes a heavier tail than the data warrant. Once NPIs are lifted, this inflated heterogeneity sustains transmission for longer and produces a notably higher projected peak than the data-generating peak, with a confidence band that separates clearly from the Gamma band. The homogeneous fit also overpredicts the small-seed epidemic substantially, producing a peak approximately four times the data-generating peak.

This asymmetry reflects the directional pattern discussed in Section 5.5.5 of the main chapter: fitting a heavier-tailed family to lighter-tailed data inflates ν more than fitting a lighter-tailed family to heavier-tailed data deflates it, and the heavier tail amplifies the forecast divergence faster than the lighter tail suppresses it. Unlike the Lognormal-baseline case (Figure 5.9), where band overlap left the misspecification practically ambiguous, here the two models produce forecasts that are visually and quantitatively distinguishable.

Summary. The two-epidemic experiments show that joint fitting sharpens identifiability when the susceptibility family is correctly specified, recovering ν and c^* with near-nominal coverage and narrow intervals. Under misspecification, the same sharpening forces the model to commit to a biased explanation: coverage collapses to 0% for ν and $\mathcal{R}_0$, and the biased estimates propagate into forecasts that overshoot the data-generating peak. The direction of the forecast error is consistent across both baselines: fitting a heavier-tailed family to lighter-tailed data produces a pessimistic forecast, whereas fitting a lighter-tailed family to heavier-tailed data produces an optimistic forecast. Multi-epidemic fitting therefore does not protect against structural misspecification; rather, it makes its consequences more visible and more consequential. produces a pessimistic forecast, while fitting a lighter-tailed family to heavier-tailed data produces an optimistic one. Multi-epidemic fitting therefore does not protect against structural misspecification but rather makes the consequences more visible and more consequential.

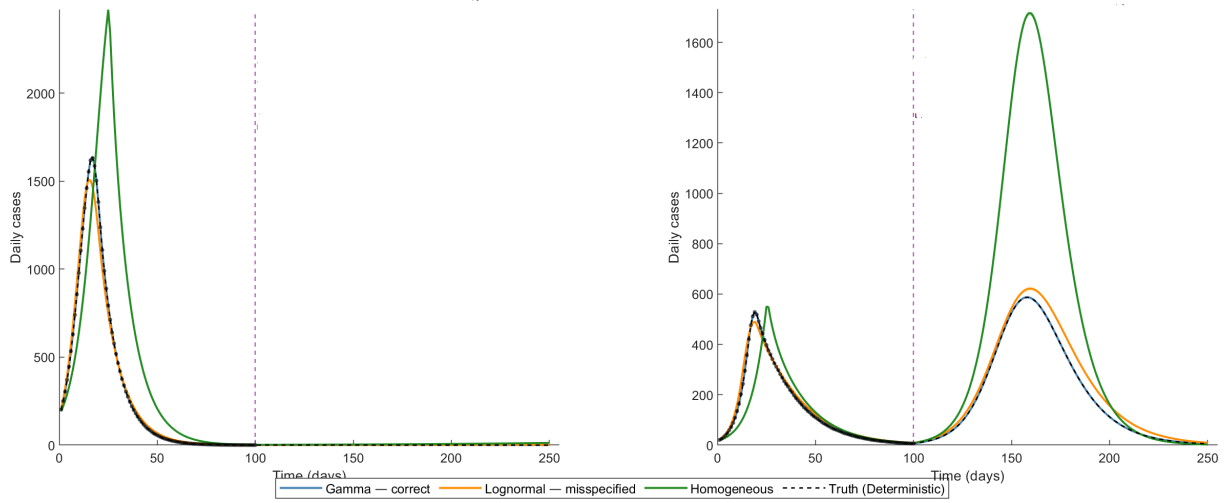


Figure F.4: Two-epidemic trajectory fits and forecasts for Gamma-generated data ($\nu = 1.414$, NPIs lifted at day 100). The black dashed line is the deterministic data-generating trajectory; shaded bands are 95% confidence envelopes across replicated fits. The correctly specified Gamma fit (blue) tracks both waves accurately. The misspecified Lognormal fit (orange) follows the Gamma fit closely in Epidemic 1 but substantially overshoots the second-wave peak in Epidemic 2, reflecting a median ν estimate of 1.781, with mean 95% Wald limits (1.768, 1.794), compared with the data-generating value 1.414. The homogeneous fit (green) also overpredicts substantially in Epidemic 2, producing a peak approximately four times the data-generating peak.

Appendix G

Supplementary Materials for Chapter 6

This appendix collects supporting material for the fits in Chapter 6.

The appendix is organised as follows:

1. Fixed inputs used in the reduced mortality model (Table G.1);
2. prior specifications for the joint hierarchical models and for comparison fits with uniform priors (Tables G.2 and G.7);
3. supporting posterior summaries for the narrower infection-to-death delay refit (Section G.1.1);
4. MCMC convergence diagnostics and posterior dependence plots for hierarchical models (Section G.2);
5. model comparison using PSIS-LOO and WAIC, together with Pareto- k reliability diagnostics (Sections G.3 and G.3.3);
6. comparison fits with uniform priors (Section G.4).

G.1 Observation-layer sensitivity and narrower-delay refit

The main observation-layer sensitivity results are reported in Section 6.7.5 of Chapter 6. That main-text subsection contains the remapping of the fitted logistic incidence under alternative infection-to-death assumptions, the corresponding England remapping summaries, and the fitted

Table G.1: Fixed inputs for the reduced SEIR mortality model used in Chapter 6.

Symbol	Meaning	Value	Source
N_j	Population size	$N_{\text{EN}} = 56 \times 10^6$, $N_{\text{SC}} = 5.5 \times 10^6$	[4]
ρ	Relative infectiousness of E	0.5	[4]
δ	Progression rate $E \rightarrow I$	$1/5.5 \text{ day}^{-1}$	[4]
γ	Recovery rate $I \rightarrow R$	$1/4 \text{ day}^{-1}$	[4]
ψ	Infection fatality ratio	0.009	[4]
L	Maximum infection-to-death lag	60 days	[14]
$(\mu_{\text{ID}}, \sigma_{\text{ID}}^2)$	Infection-to-death delay	(3.151, 0.469 ²)	[14].

and residual plots from the narrower-delay logistic refit. Those figures and summary tables are therefore not repeated here.

This appendix reports the detailed posterior summaries for the narrower-delay hierarchical logistic refit.

G.1.1 Detailed posterior summaries for the narrower-delay logistic refit

The hierarchical logistic model was refitted using the narrower infection-to-death delay

$$\log(D) \sim \mathcal{N}(3.151, 0.25^2),$$

whereas the baseline Chapter 6 analysis used $\log(D) \sim \mathcal{N}(3.151, 0.469^2)$. Tables G.3 and G.4 give the corresponding population-level and country-level posterior summaries.

Compared with the baseline logistic fit in Chapter 6, the narrower-delay refit gives a similar population-level reproduction number but moves the inferred contact floor and shared heterogeneity upward. At the country level, both England and Scotland move toward higher fitted transmission and weaker maximal contact reduction. These detailed summaries support the main-text conclusion that the observation layer is a substantive part of the empirical inference, rather than only a technical conversion from infections to deaths.

G.2 MCMC Convergence Diagnostics

All reported models were fitted using Hamiltonian Monte Carlo (HMC) with 4 chains and 2500 iterations per chain (1000 warmup). The figures in this section summarise sampling behaviour for selected common parameters for the three joint hierarchical NPI specifications (piecewise linear, logistic transition, and stringency-driven). We focus on trace plots for $\mathcal{R}_0$, ν , and c^* , which are the core quantities discussed in the main text.

Table G.2: Weakly-informed prior distributions for the hierarchical heterogeneous-susceptibility models. Half-normal priors are truncated at zero. Population-level hyperpriors generate country-specific parameters through a non-centred parameterisation.

Parameter	NPI form	Prior	Natural-scale interpretation
<i>Transmission</i>			
$\mu_{\log R_0}$	All	$\mathcal{N}(\log 3, 0.5)$	Population median of $\mathcal{R}_0$ centred at 3
$\sigma_{\log R_0}$	All	Half- $\mathcal{N}(0, 0.5)$	Between-country spread on log scale
<i>NPI parameters</i>			
$\mu_{\log w}$	Linear	$\mathcal{N}(\log 20, 1.0)$	Population median of w centred at 20 days
$\sigma_{\log w}$	Linear	Half- $\mathcal{N}(0, 1.0)$	Between-country spread on log scale
μ_{t_m}	Logistic	$\mathcal{N}(40, 5)$	Population midpoint centred at day 40
σ_{t_m}	Logistic	Half- $\mathcal{N}(0, 5)$	Between-country spread in days
$\mu_{\log k}$	Logistic	$\mathcal{N}(\log 0.35, 0.5)$	Population median of k centred at 0.35
$\sigma_{\log k}$	Logistic	Half- $\mathcal{N}(0, 0.5)$	Between-country spread on log scale
$\mu_{\text{logit } c^*}$	All	$\mathcal{N}(\text{logit}(0.33), 0.7)$	Population centre of c_j^* has median 0.33
$\sigma_{\text{logit } c^*}$	All	Half- $\mathcal{N}(0, 0.7)$	Between-country spread on logit scale
<i>Heterogeneity & dispersion</i>			
$CV_{\text{shared}} (= \nu)$	All	$\mathcal{N}(1.6, 0.35)$	Shared heterogeneity parameter on natural scale
ϕ_j	All	Log- $\mathcal{N}(\log 15, 0.6)$	Country-specific dispersion; median 15
<i>Initial conditions</i>			
$\log E_{0,\text{EN}}$	All	$\mathcal{N}(\log 100, 0.1)$	England prior median = 100
$\log E_{0,\text{SC}}$	All	$\mathcal{N}(\log 10, 0.1)$	Scotland prior median = 10
$\log I_{0,\text{EN}}$	All	$\mathcal{N}(\log(100/2.5), 0.1)$	England prior median = 40
$\log I_{0,\text{SC}}$	All	$\mathcal{N}(\log(10/2.5), 0.1)$	Scotland prior median = 4

Note: For log-scale priors, the natural-scale median is $\exp(\mu)$; for logit-scale priors, it is $\text{logit}^{-1}(\mu)$; and for priors defined directly on the natural scale, it is μ .

Table G.3: Population-level posterior summaries for the hierarchical logistic model fitted with infection-to-death delay $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$. Entries are posterior medians with posterior SD in parentheses and central 95% credible intervals in brackets.

Parameter	Median	SD	2.5%	97.5%
$R_{0,\text{pop}}$	3.03	0.612	2.10	4.48
k_{pop}	0.533	0.184	0.257	1.00
$\bar{t}_m$	45.8	0.448	44.9	46.7
c_{pop}^*	0.311	0.053	0.211	0.439
ν	1.84	0.260	1.28	2.30

G.2.1 Hierarchical Model Diagnostics

Figures G.1–G.3 report MCMC trace plots for key parameters for the three joint hierarchical NPI specifications (piecewise linear, logistic, and stringency-driven) with posterior summaries in section 6.7.3. The traces are used to verify stationarity and adequate mixing across the four chains and to confirm that posterior exploration is stable for the parameters used in the main-text comparisons.

In addition to convergence, we report posterior dependence diagnostics that are directly relevant to the identifiability themes of the thesis. Figures G.4–G.5 plot posterior draws of the

Table G.4: Country-level posterior summaries for the hierarchical logistic model fitted with infection-to-death delay $\log(D) \sim \mathcal{N}(3.151, 0.25^2)$. Entries are posterior medians with posterior SD and central 95% credible intervals. England rows are followed by Scotland rows.

Parameter	Median	SD	2.5%	97.5%
England				
$R_{0,EN}$	3.20	0.0367	3.13	3.27
c_{EN}^*	0.295	0.0163	0.266	0.329
ϕ_{EN}	335	111	188	618
$t_{m,EN}$	44.2	0.275	43.7	44.8
k_{EN}	0.598	0.255	0.420	1.24
Scotland				
$R_{0,SC}$	2.86	0.0463	2.78	2.96
c_{SC}^*	0.324	0.0167	0.291	0.356
ϕ_{SC}	26.5	9.09	15.1	50.0
$t_{m,SC}$	47.4	0.842	45.7	49.0
k_{SC}	0.553	0.399	0.229	1.56

ϕ_j is the Negative-Binomial dispersion (NB2 parameterisation) for country j . For the logistic specification, $(t_{m,j}, k_j)$ describe the midpoint and steepness of contact reduction. The shared susceptibility-heterogeneity parameter ν is reported in the final row for completeness.

shared heterogeneity parameter ν against the contact reduction c^* for each country and each NPI specification.

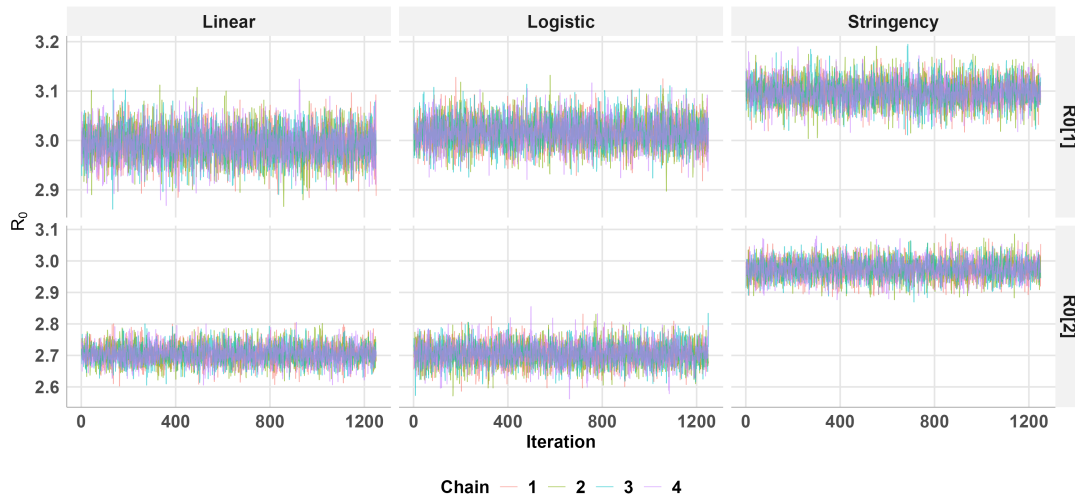


Figure G.1: MCMC trace plots for the basic reproduction number $\mathcal{R}_0$ across the three joint hierarchical NPI models: piecewise linear, logistic, and stringency-driven. The four chains mix well and remain stable over the post-warmup iterations, with no visible long-term drift or persistent separation between chains. This supports satisfactory convergence for the transmissibility parameter across all three NPI specifications.

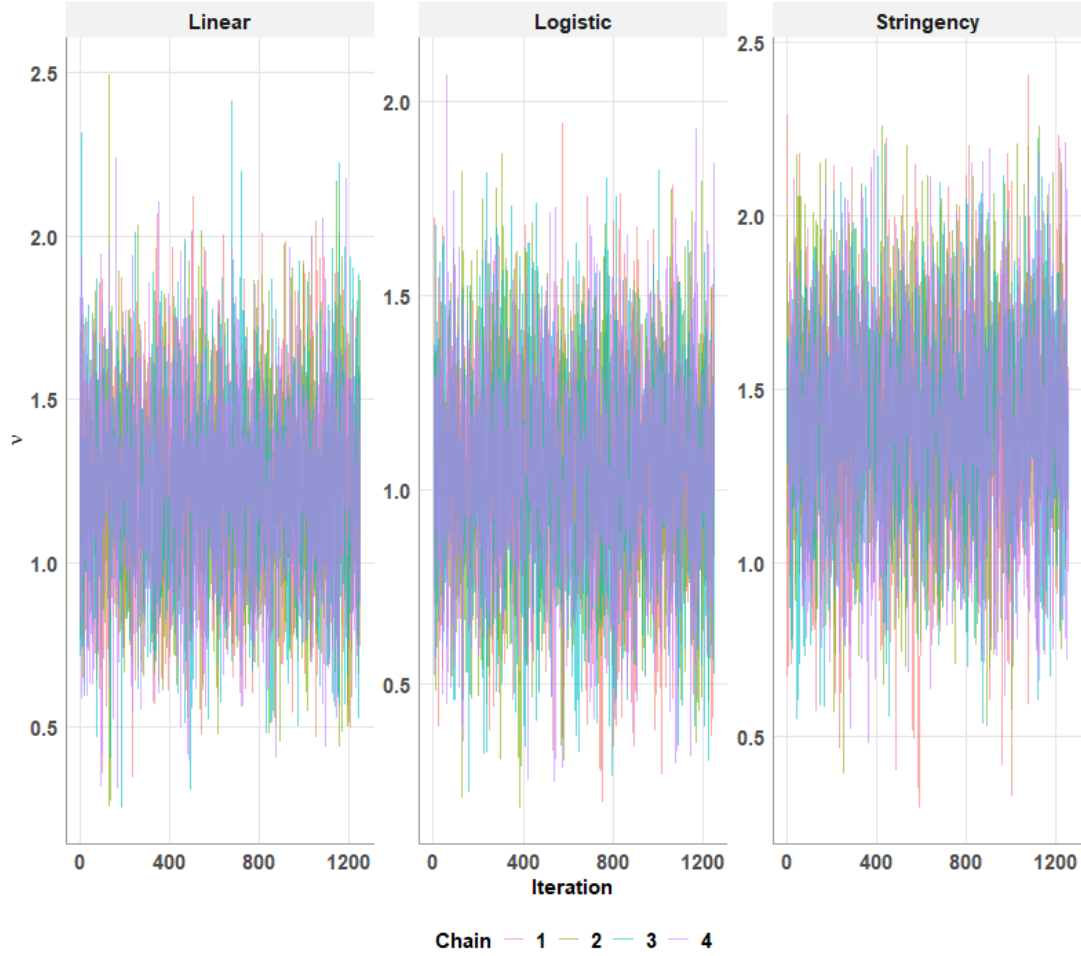


Figure G.2: MCMC trace plots for the shared susceptibility-heterogeneity parameter ν across the three joint hierarchical NPI models: piecewise linear, logistic, and stringency-driven. The four chains show stable mixing and no persistent non-stationary behaviour. The stringency-driven model explores a somewhat higher region of the parameter space, reflecting its larger posterior values of ν in the main results, but overall the traces support satisfactory convergence across specifications with wider intervals for the stringency index.

G.3 Hierarchical Model Comparison and Selection

We compare the three joint hierarchical NPI specifications (piecewise linear ramp, logistic transition, and stringency-driven) using Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO) and the widely applicable information criterion (WAIC), following [106]. Both criteria are reported on the expected log pointwise predictive density scale (elpd), where *higher* elpd indicates better expected out-of-sample predictive accuracy. Each observation corresponds to a country–day outcome in the joint likelihood, so differences in elpd quantify expected predictive performance for held-out country–day observations.

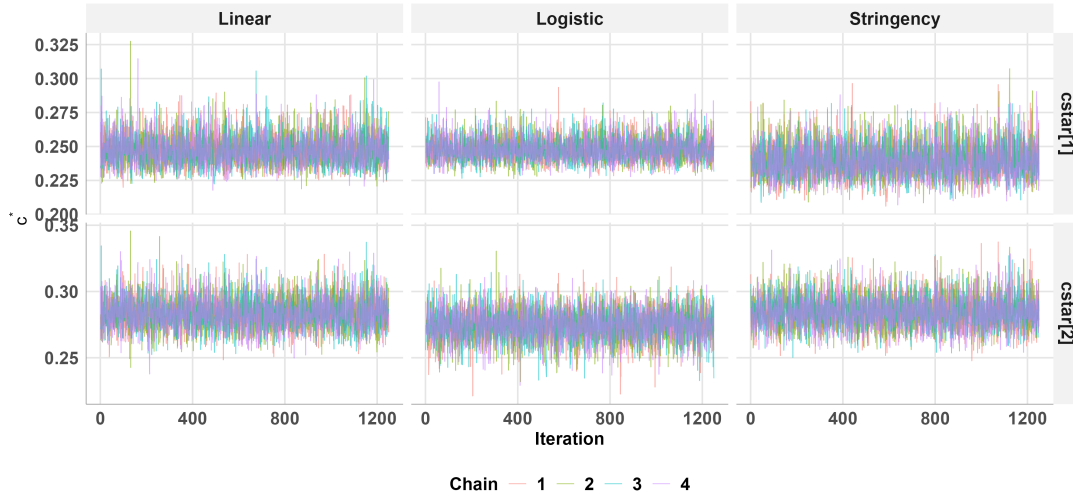


Figure G.3: MCMC trace plots for the contact reduction parameter (c^*). The sampling chains for all three models achieve stationarity and converge on a similar median estimate, indicating that contacts were reduced to roughly 26% to 28% of their pre-lockdown baseline. However, similar to the $\mathcal{R}_0$ and ν parameters, the stringency model's chains have a marginally wider dispersion and slightly different mixing characteristics during exploration. This subtle divergence highlights the increased uncertainty the sampler encounters when forcing an exact, data-driven step-function (OxCGRT) to explain the smooth, observed decline in mortality.

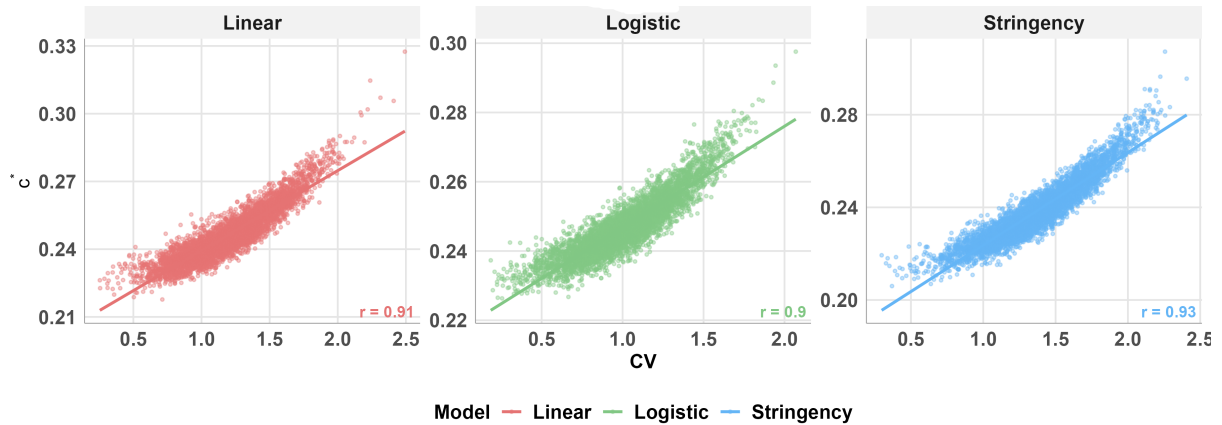


Figure G.4: Posterior dependence between the shared heterogeneity parameter ν and the contact-reduction parameter c^* for England, shown separately for the piecewise linear, logistic, and stringency-driven models. Points are posterior draws, and the fitted line is a least-squares regression on the plotted draws. The correlation coefficients reported in the panels are all positive, indicating the expected compensation pattern in which higher inferred heterogeneity is associated with weaker inferred maximal contact reduction (larger c^*).

G.3.1 Three-Way Comparison

Table G.5 reports the three-way comparison across the hierarchical NPI models. The best-performing model is taken as the reference and assigned $\Delta\text{elpd} = 0$; the remaining rows report differences relative to that reference, with standard errors (s.e.) of the differences. Both PSIS-LOO and WAIC select the **logistic** specification as the best-performing model, with the piecewise

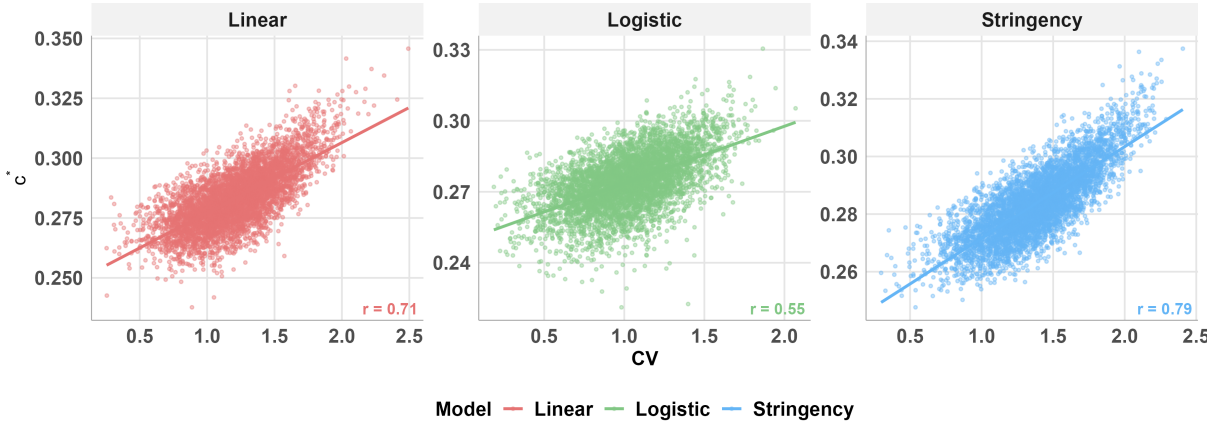


Figure G.5: Posterior dependence between CV and the intervention c^* for **Scotland**, shown for each NPI specification. Points are posterior draws and the fitted line is a least-squares regression on the plotted draws; the reported correlation coefficient summarises the strength of association. Scotland’s fit has reduced correlation compared to England’s (Figure G.4) and therefore more in agreement with the gain we saw in the synthetic data study.

linear model lower by $\Delta\text{elpd} = -18.5$ (s.e. 2.9) and the stringency-driven model lower by $\Delta\text{elpd} = -31.0$ (s.e. 5.9).

Table G.5: PSIS-LOO and WAIC comparison across the three joint hierarchical NPI models. Entries are differences in elpd relative to the best model ($\Delta\text{elpd} = 0$), with standard errors of the differences. Higher elpd indicates better expected out-of-sample predictive accuracy [106].

Model	PSIS-LOO		WAIC	
	Δelpd	s.e.	Δelpd	s.e.
Logistic	0.0	0.0	0.0	0.0
Linear	-18.5	2.9	-18.5	2.9
Stringency	-31.0	5.9	-31.0	5.9

G.3.2 Hierarchical models pairwise Comparisons

For completeness, we also report pairwise model comparisons (Table G.6). These are consistent with the three-way ranking: logistic is preferred over both alternatives, and the linear model is preferred over the stringency-driven model. In particular, relative to the linear ramp, the stringency-driven specification has lower predictive accuracy ($\Delta\text{elpd} = -12.5$, s.e. 4.6).

G.3.3 Hierarchical Model Comparison Reliability

We assess the reliability of PSIS-LOO using Pareto- k diagnostics [106]. As a practical rule-of-thumb, $k < 0.5$ indicates stable importance sampling; values in $[0.5, 0.7]$ warrant caution; and $k \gtrsim 0.7$ suggests that the PSIS approximation may be unreliable for the affected observations,

Table G.6: Pairwise PSIS-LOO and WAIC comparisons among the three joint hierarchical NPI models. For each pair, the best model is the reference ($\Delta\text{elpd} = 0$). Higher elpd indicates better predictive accuracy [106].

Comparison	PSIS-LOO		WAIC	
	Δelpd	s.e.	Δelpd	s.e.
Logistic vs Linear (Linear–Logistic)	−18.5	2.9	−18.5	2.9
Logistic vs Stringency (Stringency–Logistic)	−31.0	5.9	−31.0	5.9
Linear vs Stringency (Stringency–Linear)	−12.5	4.6	−12.5	4.6

motivating exact LOO for those points or K -fold cross-validation.

Figure G.6 shows the distribution of Pareto- k values for the three hierarchical NPI models. The mass of the distributions lies well below the warning threshold, indicating that the PSIS-LOO approximation is stable for this application and that the model-ranking conclusions in Tables G.5–G.6 are unlikely to be driven by unstable importance sampling.

Table G.7: Prior distributions for centred models for the three NPI specifications. All estimated parameters have uniform priors implemented via parameter bounds in Stan. Countries are fitted with no information sharing.

Parameter	Linear ramp	Logistic	Stringency
$R_{0,j}$	$\mathcal{U}(2, 8)$	$\mathcal{U}(2, 8)$	$\mathcal{U}(2, 8)$
c_j^*	$\mathcal{U}(0.05, 0.9)$	$\mathcal{U}(0.05, 0.9)$	$\mathcal{U}(0.05, 0.9)$
$t_{0,j}$	(fixed / implied)	–	–
k_j	–	$\mathcal{U}(0.01, 2.0)$	–
$t_{m,j}$	–	$\mathcal{U}(1, T)$	–
ν	$\mathcal{U}(0.01, 3)$	$\mathcal{U}(0.01, 3)$	$\mathcal{U}(0.01, 3)$
ϕ_j	$\mathcal{U}(10, 1000)$	$\mathcal{U}(10, 1000)$	$\mathcal{U}(10, 1000)$
$I_{0,j}$	$\mathcal{U}(10, 1000)$	$\mathcal{U}(10, 1000)$	$\mathcal{U}(10, 10000)$

T denotes the length of the analysis window (number of daily observations) used to bound the logistic midpoint, i.e. $t_{m,j} \in [1, T]$. The stringency-driven specification uses the observed OxCGRT-derived covariate $u_{j,t} \in [0, 1]$ to define $c(t)$ and therefore has no additional timing/shape parameters beyond c_j^* .

G.4 Fits with Uniform Priors

This section reports *fits* obtained under the uniform priors in Table G.7. The posterior summaries are discussed in Section 6.8, Tables 6.8 and 6.9.

G.4.1 Model Fit and Residual Diagnostics

Figures G.9–G.10 show posterior predictive fits (median and 95% pointwise intervals) together with the implied contact modifier $c(t)$ for each NPI specification. Figure G.7 complements these trajectory plots by showing residuals over time.

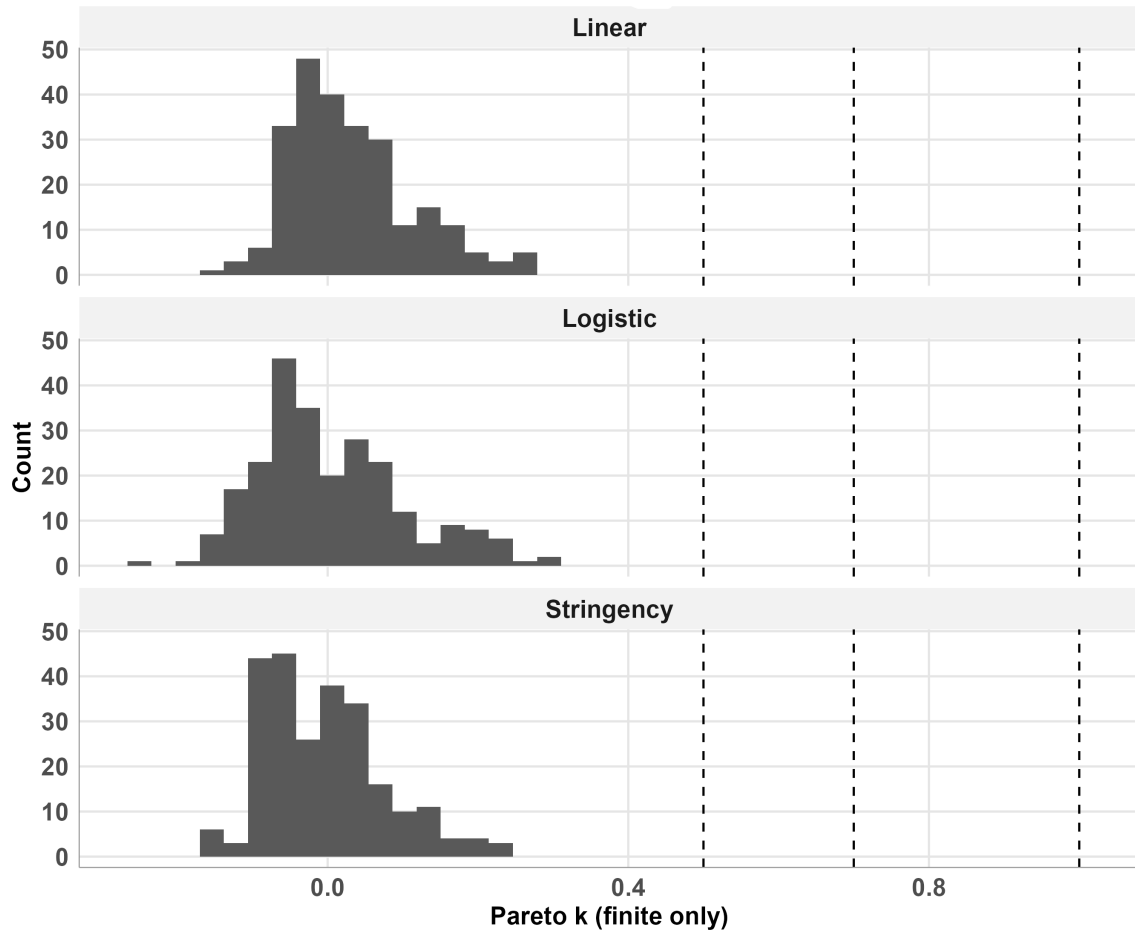


Figure G.6: Distribution of Pareto- k values for the three joint hierarchical NPI models: piecewise linear, logistic, and stringency-driven. Dashed vertical lines mark common diagnostic thresholds ($k = 0.5$ good, $k = 0.7$ caution, $k = 1.0$ poor). Most values lie below the warning threshold, indicating that the PSIS-LOO approximation is reliable for the model comparisons reported in Section G.3.

G.4.2 Posterior distribution and correlation plots comparisons

Figures G.8-G.12 compare marginal posterior distributions across three NPI specifications for uniform prior fit for $\mathcal{R}_0$, ν and c^* . Correlation plots are also compared in Figures G.13-G.14.

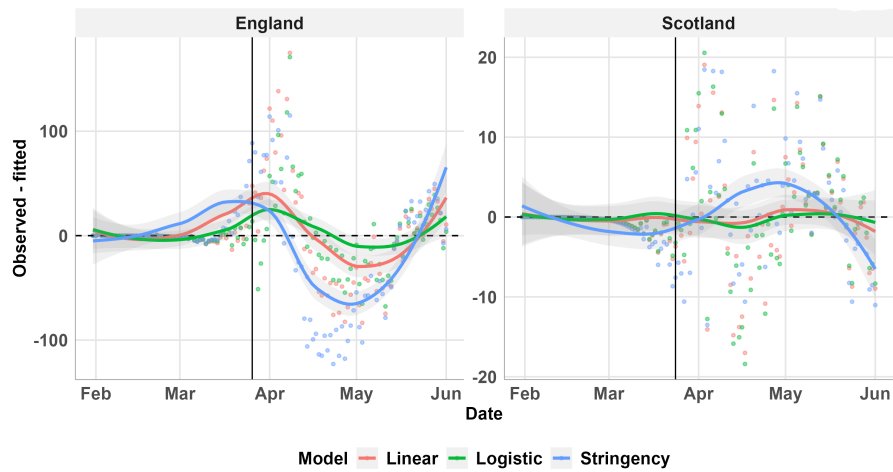


Figure G.7: Residuals for the three NPI specifications for uniform priors, shown for England (left) and Scotland (right). Points are daily residuals; smooth curves summarise residual trends over time for each model. The vertical line denotes the official lockdown date.

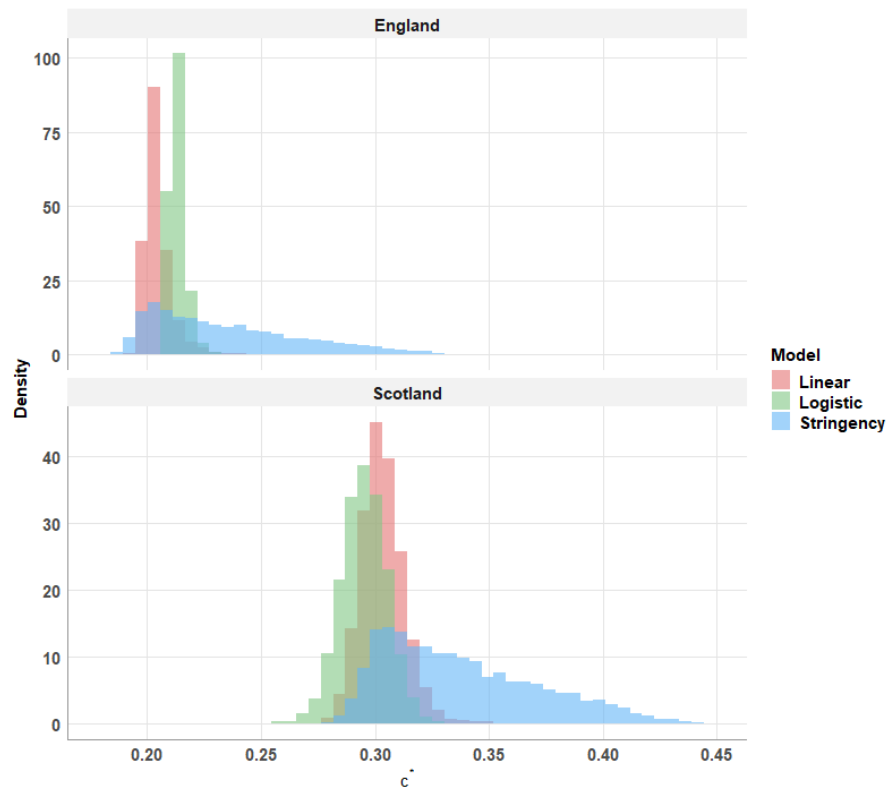


Figure G.8: Posterior comparison of the minimum contact level c^* for uniform priors, shown for England (top) and Scotland (bottom). Smaller values of c^* correspond to stronger maximal contact reduction. Within each country, the linear and logistic models give fairly similar posterior locations, whereas the stringency-driven model produces a broader and more shifted distribution, especially in Scotland. This agrees with the country-level summaries in Table 6.9.

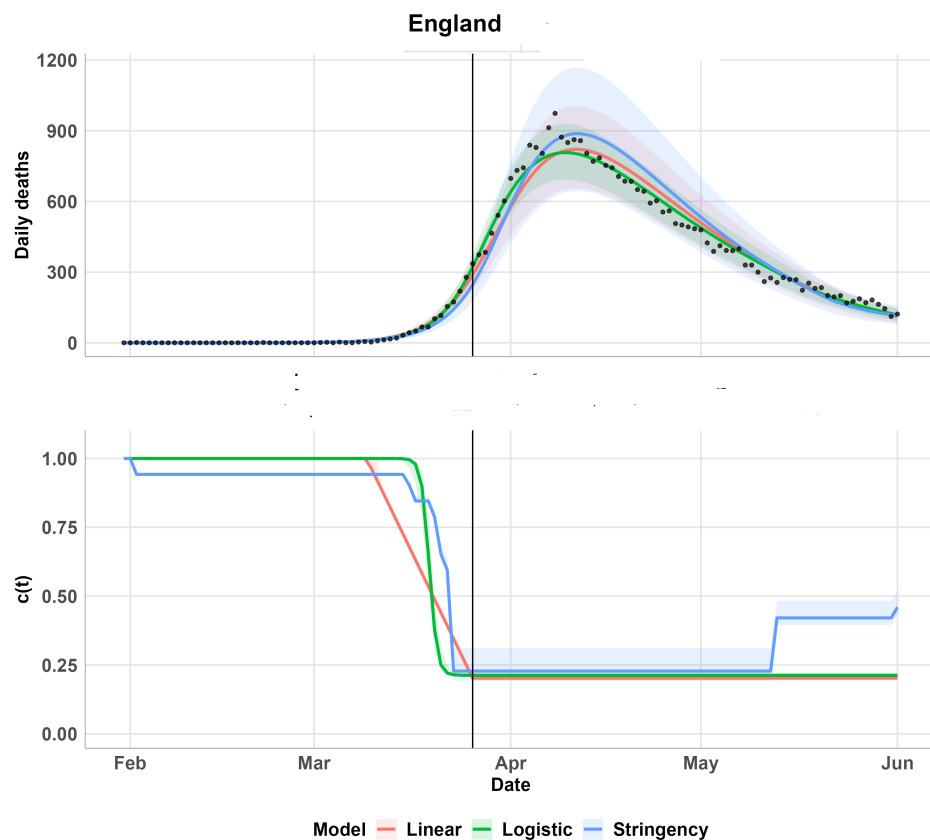


Figure G.9: Comparison fits for England with uniform priors. **Top:** Posterior predictive trajectories for daily deaths (posterior median and 95% pointwise interval) corresponding to the piecewise linear (red), logistic (green), and stringency-driven (blue) NPI specifications. **Bottom:** The corresponding implied contact modifier $c(t)$ for each specification. The vertical line indicates the official lockdown date.

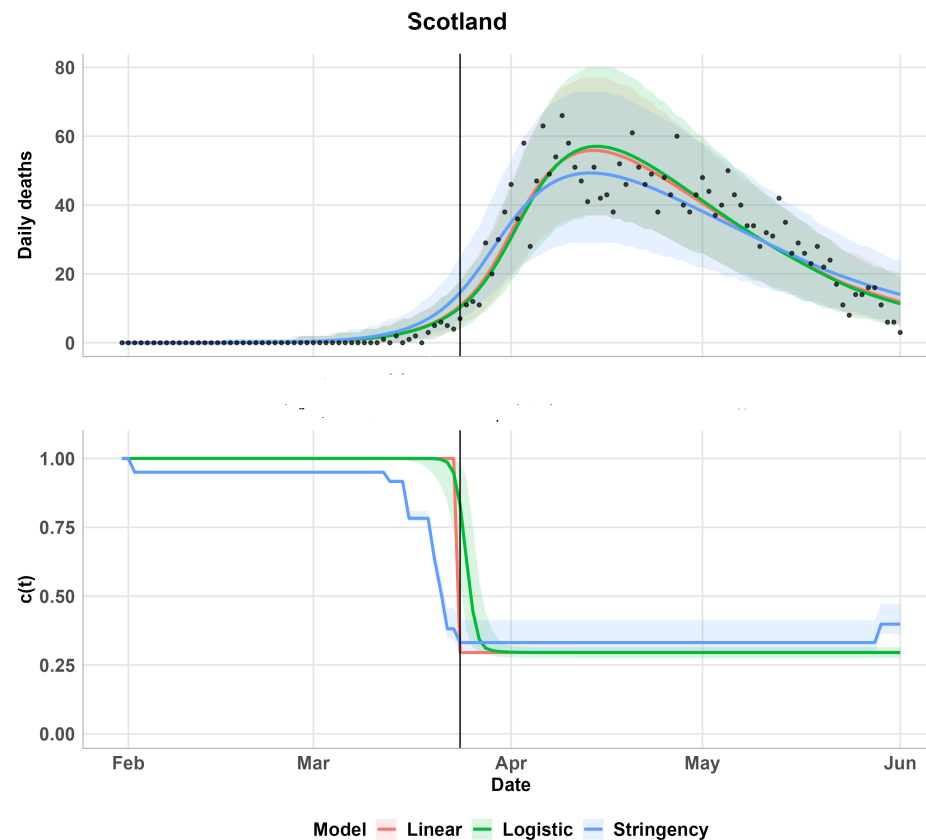


Figure G.10: Comparison fits for Scotland for uniform priors. **Top:** Posterior predictive trajectories for daily deaths (posterior median and 95% pointwise interval) for the piecewise linear (red), logistic (green), and stringency-driven (blue) NPI specifications. **Bottom:** The corresponding implied contact modifier $c(t)$ for each specification. The vertical line indicates the official lockdown date.

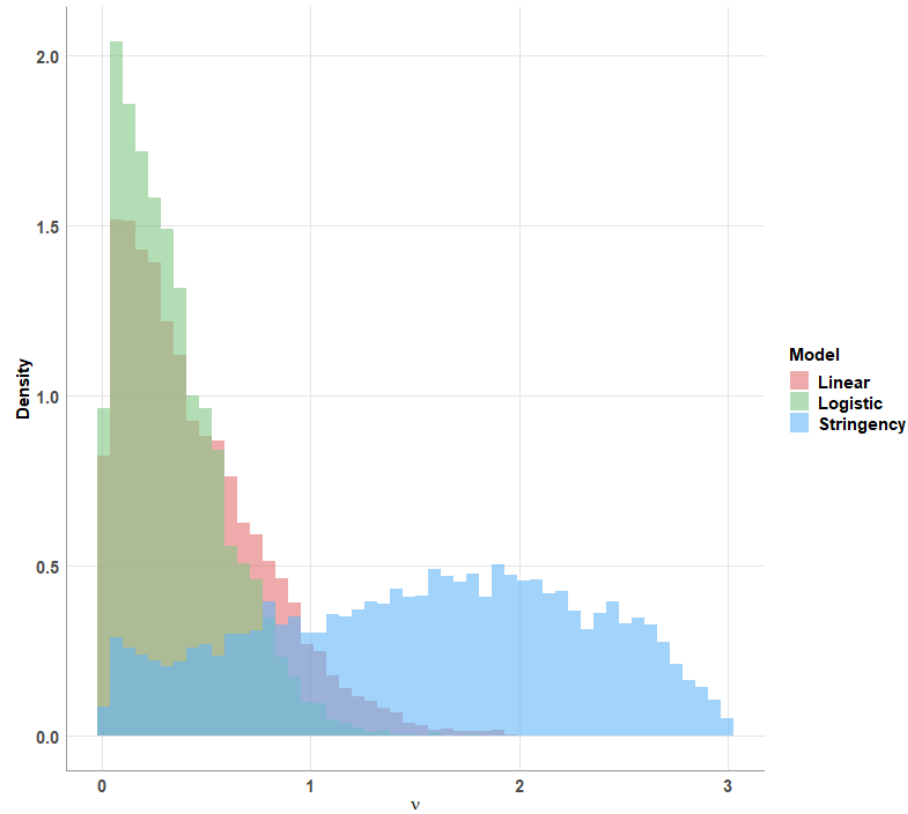


Figure G.11: Posterior comparison of the shared susceptibility-heterogeneity parameter ν for uniform priors. The linear and logistic models place most posterior mass near low values of ν , close to the homogeneous limit, whereas the stringency-driven model shifts posterior mass toward across all range of the prior range of CV (Table G.7) showing no information from the data on coefficient of variation.

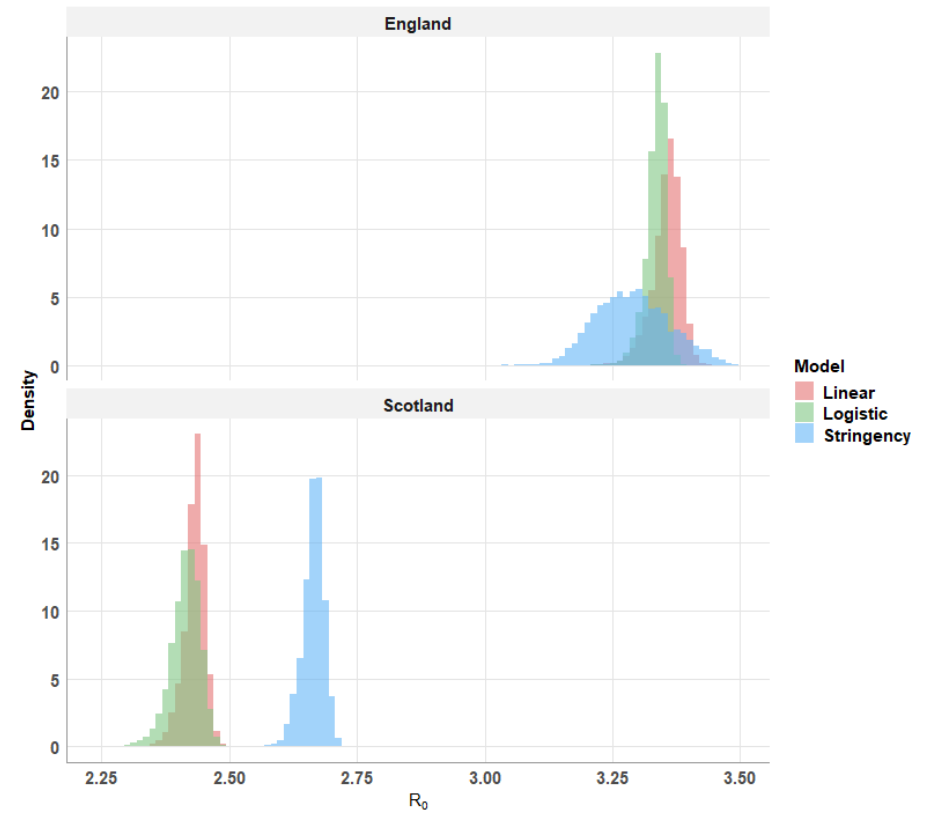


Figure G.12: Uniform-prior posterior comparison of $\mathcal{R}_0$. Posterior comparison of the country-specific basic reproduction number $\mathcal{R}_0$ with uniform priors, shown for England (top figure) and Scotland (bottom figure). Within each country, the linear and logistic models give similar posterior distributions, while the stringency-driven model shifts $\mathcal{R}_0$ upward, especially for Scotland. This matches the country-level summaries in Table 6.9, and illustrates that with broad priors the stringency specification reallocates explanatory weight toward higher baseline transmissibility and stronger heterogeneity.

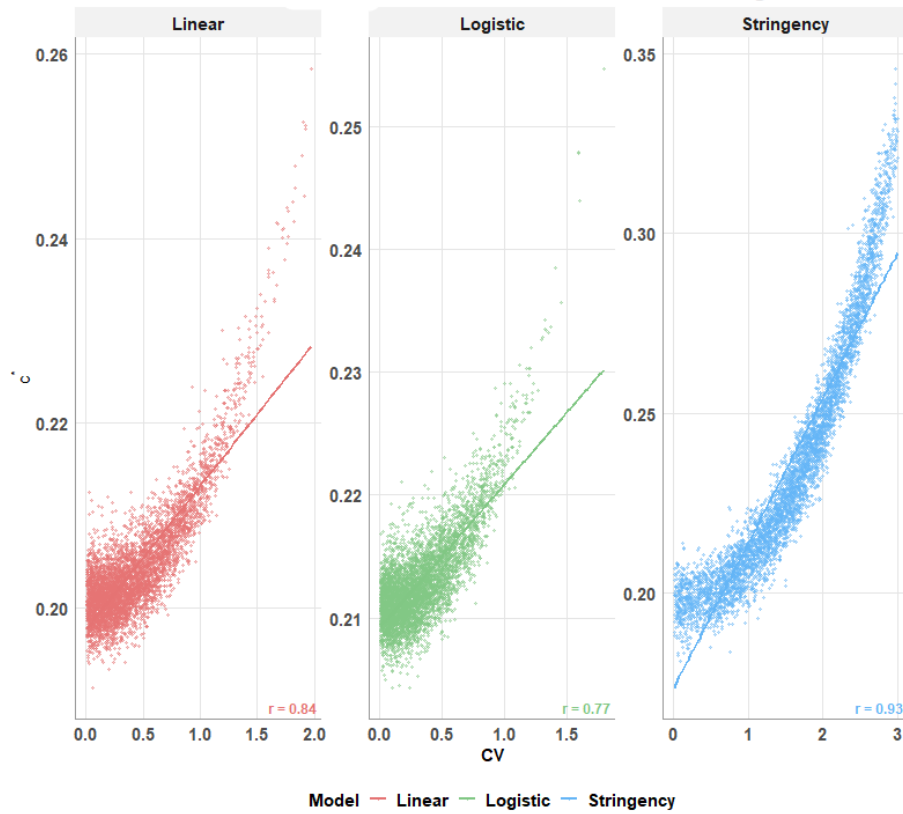


Figure G.13: Uniform-prior posterior correlation between ν and c^* in England. Posterior dependence between the shared heterogeneity parameter ν and the contact-reduction parameter c^* for England in the comparison fits with uniform priors, shown for the piecewise linear, logistic, and stringency-driven models. Points are posterior draws, and the fitted line is a least-squares regression on the plotted draws. The panel-wise correlation coefficients are positive, indicating the same ν - c^* compensation pattern seen in the hierarchical analysis, but with weaker regularisation and greater sensitivity to the assumed NPI form.

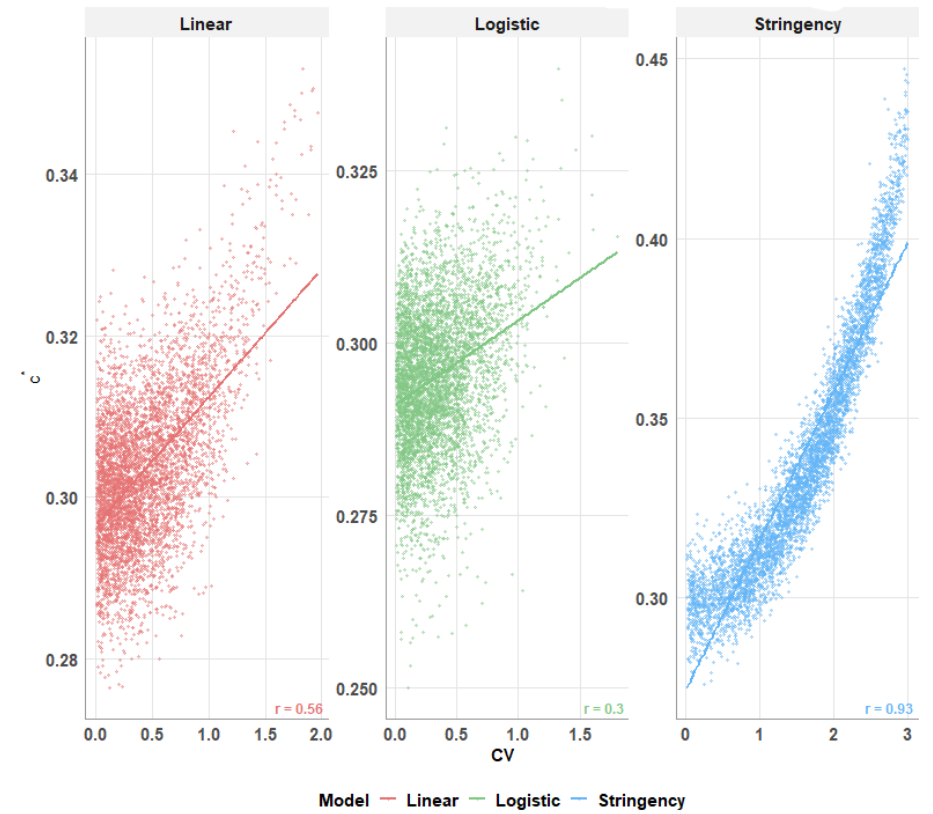


Figure G.14: Uniform-prior posterior correlation between ν and c^* in Scotland. Posterior dependence between the shared heterogeneity parameter ν and the contact-reduction parameter c^* for Scotland in the comparison fits with uniform priors, shown for the piecewise linear, logistic, and stringency-driven models. Points are posterior draws, and the fitted line is a least-squares regression on the plotted draws. The panel-wise correlation coefficients remain positive, but are generally weaker than for England, consistent with the reduced ν - c^* compensation seen for Scotland throughout the chapter.

Bibliography

- [1] Ibrahim Mohammed, Chris Robertson, and M. Gabriela M. Gomes. On the simultaneous inference of susceptibility distributions and intervention effects from epidemic curves. *Epidemics*, 55:100911, June 2026.
- [2] M. Gabriela M. Gomes, Ibrahim Mohammed, and Chris Robertson. M. Gabriela M. Gomes, Ibrahim Mohammed and Chris Robertson’s contribution to the Discussion of “Some statistical aspects of the COVID-19 response” by Wood et al. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 189(1):43–45, January 2026.
- [3] Ibrahim Mohammed, Chris Robertson, and M. Gabriela M. Gomes. An exploration into how susceptibility distribution misspecifications impact epidemic forecasting. *arXiv preprint arXiv:2606.08693*, 2026.
- [4] M. Gabriela M. Gomes, Marcelo U. Ferreira, Rodrigo M. Corder, Jessica G. King, Caetano Souto-Maior, Carlos Penha-Gonçalves, Guilherme Gonçalves, Maria Chikina, Wesley Pegden, and Ricardo Aguas. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *Journal of Theoretical Biology*, 540:111063, 2022.
- [5] Adam Kleczkowski, Andy Hoyle, and Paul McMenemy. One model to rule them all? modelling approaches across OneHealth for human, animal and plant epidemics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1775):20180255, 2019.
- [6] Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.
- [7] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.
- [8] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

- [9] W. H. Hamer. The Milroy Lectures on epidemic disease in England — the evidence of variability and of persistency of type. *The Lancet*, 167(4307):733–739, 1906.
- [10] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A*, 115:700–721, 1927.
- [11] Ronald Ross. *The Prevention of Malaria*. John Murray, London, 2nd edition, 1911.
- [12] Emilia Vynnycky and Richard G. White. *An Introduction to Infectious Disease Modelling*. Oxford University Press, 2010.
- [13] Kimberlyn Roosa and Gerardo Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*, 16(1):1, 2019.
- [14] Simon N. Wood, Ernst C. Wit, Paul M. McKeigue, Danshu Hu, Beth Flood, Lauren Corcoran, and Thea Abou Jawad. Some statistical aspects of the COVID-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 189(1):1–31, 2026.
- [15] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan Anton Jacob Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.
- [16] Pauline Van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1–2):29–48, 2002.
- [17] Paul Fine, Ken Eames, and David L. Heymann. “Herd Immunity”: A Rough Guide. *Clinical Infectious Diseases*, 52(7):911–916, 2011.
- [18] Stavros Busenberg and Carlos Castillo-Chavez. A general solution of the problem of mixing of subpopulations and its application to risk- and age-structured epidemic models for the spread of AIDS. *IMA Journal of Mathematics Applied in Medicine & Biology*, 8(1):1–29, 1991.
- [19] P. Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157–172, 1983.
- [20] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.

- [21] Adam Kleczkowski and Bryan T. Grenfell. Mean-field-type equations for spread of epidemics: the ‘small world’ model. *Physica A: Statistical Mechanics and its Applications*, 274(1–2):355–360, 1999.
- [22] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.
- [23] Wei-min Liu, Herbert W. Hethcote, and Simon A. Levin. Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of Mathematical Biology*, 25(4):359–380, 1987.
- [24] Michael E. Hochberg. Non-linear transmission rates and the dynamics of infectious disease. *Journal of Theoretical Biology*, 153(3):301–321, 1991.
- [25] Jonas Neipel, Jonathan Bauermann, Stefano Bo, Tyler Harmon, and Frank Jülicher. Power-law population heterogeneity governs epidemic waves. *PLOS ONE*, 15(10):e0239678, 2020.
- [26] Hamish McCallum, Nigel Barlow, and Jim Hone. How should pathogen transmission be modelled? *Trends in Ecology & Evolution*, 16(6):295–300, 2001.
- [27] M. J. Keeling and B. T. Grenfell. Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Mathematical Biosciences*, 147(2):207–226, 1998.
- [28] Helen J. Wearing, Pejman Rohani, and Matt J. Keeling. Appropriate models for the management of infectious diseases. *PLOS Medicine*, 2(7):e174, 2005.
- [29] Neil M. Ferguson, Derek A. T. Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sophon Iamsirithaworn, and Donald S. Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, 2005.
- [30] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.
- [31] Neil M. Ferguson, Derek A. T. Cummings, Christophe Fraser, James C. Cajka, Philip C. Cooley, and Donald S. Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.

- [32] Özden O. Dalgıç, Osman Y. Özaltın, William A. Ciccotelli, and Fatih S. Erenay. Deriving effective vaccine allocation strategies for pandemic influenza: Comparison of an agent-based simulation and a compartmental model. *PLOS ONE*, 12(2):e0172261, 2017.
- [33] Kiesha Prem, Alex R. Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13(9):e1005697, 2017.
- [34] Hamdy Youssef, Najat Alghamdi, Magdy A. Ezzat, Alaa A. El-Bary, and Ahmed M. Shawky. Study on the SEIQR model and applying the epidemiological rates of COVID-19 epidemic spread in Saudi Arabia. *Infectious Disease Modelling*, 6:678–692, 2021.
- [35] Chuang Xu, Yongzhen Pei, Shengqiang Liu, and Jinzhi Lei. Effectiveness of non-pharmaceutical interventions against local transmission of COVID-19: An individual-based modelling study. *Infectious Disease Modelling*, 6:848–858, 2021.
- [36] Xia Wang, Hulin Wu, and Sanyi Tang. Assessing age-specific vaccination strategies and post-vaccination reopening policies for COVID-19 control using SEIR modeling approach. *Bulletin of Mathematical Biology*, 84(10):108, 2022.
- [37] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505):846–849, 2020.
- [38] Calistus N. Ngonghala, Hemaho B. Taboe, Salman Safdar, and Abba B. Gumel. Unraveling the dynamics of the Omicron and Delta variants of the 2019 coronavirus in the presence of vaccination, mask usage, and antiviral treatment. *Applied Mathematical Modelling*, 114:447–465, 2023.
- [39] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4):529–538, 2021.
- [40] Hisashi Inaba. *Age-Structured Population Dynamics in Demography and Epidemiology*. Springer, Singapore, 2017.
- [41] István Z. Kiss, Joel C. Miller, and Péter L. Simon. *Mathematics of Epidemics on Networks: From Exact to Approximate Models*. Springer, Cham, 2017.

- [42] Gilbert G. Walter and Martha Contreras. *Compartmental Modeling with Networks*. Springer Science & Business Media, 1999.
- [43] Christopher Rose, Andrew J. Medford, C. Franklin Goldsmith, Tejs Vegge, Joshua S. Weitz, and Andrew A. Peterson. Heterogeneity in susceptibility dictates the order of epidemic models. *Journal of Theoretical Biology*, 528:110839, 2021.
- [44] Antonio Montalbán, Rodrigo M. Corder, and M. Gabriela M. Gomes. Herd immunity under individual variation and reinfection. *Journal of Mathematical Biology*, 85(1):2, 2022.
- [45] John J. Gart. The mathematical analysis of an epidemic with two kinds of susceptibles. *Biometrics*, 24(3):557–566, 1968.
- [46] John J. Gart. The statistical analysis of chain-binomial epidemic models with several kinds of susceptibles. *Biometrics*, 28(4):921–930, 1972.
- [47] Frank Ball. Deterministic and stochastic epidemics with several kinds of susceptibles. *Advances in Applied Probability*, 17(1):1–22, 1985.
- [48] James M. Hyman and Jia Li. Differential susceptibility epidemic models. *Journal of Mathematical Biology*, 50(6):626–644, 2005.
- [49] Xiao-Bing Zhang, Hai-Feng Huo, Xiao-Ke Sun, and Qiang Fu. The differential susceptibility SIR epidemic model with time delay and pulse vaccination. *Journal of Applied Mathematics and Computing*, 34(1):287–298, 2010.
- [50] Xiao-Bing Zhang, Hai-Feng Huo, Xiao-Ke Sun, and Qiang Fu. The differential susceptibility SIR epidemic model with stage structure and pulse vaccination. *Nonlinear Analysis: Real World Applications*, 11(4):2634–2646, 2010.
- [51] Yukihiro Nakata and Ryosuke Omori. Epidemic dynamics with a time-varying susceptibility due to repeated infections. *Journal of Biological Dynamics*, 13(1):567–585, 2019.
- [52] Artem S. Novozhilov. On the spread of epidemics in a closed heterogeneous population. *Mathematical Biosciences*, 215(2):177–185, 2008.
- [53] Guy Katriel. The size of epidemics in populations with heterogeneous susceptibility. *Journal of Mathematical Biology*, 65(2):237–262, 2012.
- [54] B. Bonzi, A. A. Fall, A. Iggidr, and G. Sallet. Stability of differential susceptibility and infectivity epidemic models. *Journal of Mathematical Biology*, 62:39–64, 2011.

- [55] Vladimir M. Veliov. On the effect of population heterogeneity on dynamics of epidemic diseases. *Journal of Mathematical Biology*, 51:123–143, 2005.
- [56] F. A. B. Coutinho, E. Massad, L. F. Lopez, M. N. Burattini, C. J. Struchiner, and R. S. Azevedo-Neto. Modelling heterogeneities in individual frailties in epidemic models. *Mathematical and Computer Modelling*, 30(1):97–115, 1999.
- [57] Odd O. Aalen, Morten Valberg, Tom Grotmol, and Steinar Tretli. Understanding variation in disease risk: the elusive concept of frailty. *International Journal of Epidemiology*, 44(4):1408–1421, 2015.
- [58] Jeremy D. Harris, Esther Gallmeier, Jonathan Dushoff, Stephen J. Beckett, and Joshua S. Weitz. Infections are not alike: the effects of covariation between individual susceptibility and transmissibility on epidemic dynamics. *Journal of Theoretical Biology*, 613:112211, 2025.
- [59] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.
- [60] Jan M. Brauner, Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B. Stephenson, Gavin Leech, George Altman, Vladimir Mikulík, Alexander John Norman, Joshua Teperowski Monrad, Tamay Besiroglu, Hong Ge, Meghan A. Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, and Jan Kulveit. Inferring the effectiveness of government interventions against COVID-19. *Science*, 371(6531):eabd9338, 2021.
- [61] Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, João Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, 369(6500):eabb9789, 2020.
- [62] Gabriel McCarthy and Hana M. Dobrovolny. Determining the best mathematical model for implementation of non-pharmaceutical interventions. *Mathematical Biosciences and Engineering*, 22(3):700–724, 2025.
- [63] Koen Deforche, Jurgen Vercauteren, Viktor Müller, and Anne-Mieke Vandamme. Behavioral changes before lockdown and decreased retail and recreation mobility during lockdown

- contributed most to controlling COVID-19 in Western countries. *BMC Public Health*, 21(1):654, 2021.
- [64] Kristian Soltesz, Fredrik Gustafsson, Toomas Timpka, Joakim Jaldén, Carl Jidling, Albin Heimerson, Thomas B. Schön, Armin Spreco, Joakim Ekberg, Örjan Dahlström, Fredrik Bagge Carlson, Anna Jöud, and Bo Bernhardsson. The effect of interventions on COVID-19. *Nature*, 588(7839):E26–E28, 2020. Matters Arising.
- [65] Seth Flaxman, Swapnil Mishra, James Scott, Neil Ferguson, Axel Gandy, and Samir Bhatt. Reply to: The effect of interventions on COVID-19. *Nature*, 588(7839):E29–E32, 2020. Matters Arising.
- [66] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, 368(6498):1481–1486, 2020.
- [67] Christopher I. Jarvis, Kevin Van Zandvoort, Amy Gimma, Kiesha Prem, Petra Klepac, G. James Rubin, W. John Edmunds, and CMMID COVID-19 working group. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Medicine*, 18(1):124, 2020.
- [68] Vincent Chin, John P. A. Ioannidis, Martin A. Tanner, and Sally Cripps. Effect estimates of COVID-19 non-pharmaceutical interventions are non-robust and highly model-dependent. *Journal of Clinical Epidemiology*, 136:96–132, 2021.
- [69] Bernhard Müller, Inken Padberg, Michael Lorke, Ralph Brinks, Sally Cripps, M. Gabriela M. Gomes, Daniel Haake, and John P. A. Ioannidis. Uncertainty and inconsistency of COVID-19 non-pharmaceutical intervention effects with multiple competitive statistical models. *Scientific Reports*, 16(1):5767, 2026.
- [70] Ann Osi and Navid Ghaffar zadegan. Parameter estimation in behavioral epidemic models with endogenous societal risk-response. *PLOS Computational Biology*, 20(3):e1011992, 2024.
- [71] Nicholas G. Davies, Adam J. Kucharski, Rosalind M. Eggo, Amy Gimma, and W. John Edmunds. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020.

- [72] Ying Liu, Albert A. Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2):taaa021, 2020.
- [73] Imperial College COVID-19 Response Team. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Technical Report 9, Imperial College London, March 2020.
- [74] R. Bellman and K. J. Åström. On structural identifiability. *Mathematical Biosciences*, 7(3):329–339, 1970.
- [75] Alejandro F. Villaverde, Antonio Barreiro, and Antonis Papachristodoulou. Structural identifiability of dynamic systems biology models. *PLOS Computational Biology*, 12(10):e1005153, 2016.
- [76] Franz-Georg Wieland, Adrian L. Hauber, Marcus Rosenblatt, Christian Tönsing, and Jens Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69, 2021.
- [77] Yuganthi R. Liyanage, Omar Saucedo, Necibe Tuncer, and Gerardo Chowell. A tutorial on structural identifiability of epidemic models using StructuralIdentifiability.jl. *arXiv preprint arXiv:2505.10517*, 2025.
- [78] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.
- [79] Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, and James P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLOS Computational Biology*, 3(10):e189, 2007.
- [80] François Ollivier. *Le problème de l’identifiabilité structurelle globale: approche théorique, méthodes effectives et bornes de complexité*. PhD thesis, École Polytechnique, Palaiseau, 1990.
- [81] Ottar N. Bjørnstad. *Epidemics: Models and Data Using R*. Use R! Springer International Publishing, Cham, 2018.

- [82] Yushuf Sharker, Zaynab Diallo, Wasiur R. KhudaBukhsh, and Eben Kenah. Pairwise accelerated failure time regression models for infectious disease transmission in close-contact groups with external sources of infection. *Statistics in Medicine*, 43(27):5138–5154, 2024.
- [83] Jiansi Gao, Michael R. May, Bruce Rannala, and Brian R. Moore. Model misspecification misleads inference of the spatial dynamics of disease outbreaks. *Proceedings of the National Academy of Sciences*, 120(11):e2213913120, 2023.
- [84] Alessandro Margheri, Carlota Rebelo, and M. Gabriela M. Gomes. Heterogeneity in disease risk induces falling vaccine protection with rising disease incidence. *Dynamical Systems*, 32(1):148–163, 2017.
- [85] M. Elizabeth Halloran, Ira M. Longini, Jr., and Claudio J. Struchiner. Estimability and interpretation of vaccine efficacy using frailty mixing models. *American Journal of Epidemiology*, 144(1):83–97, July 1996.
- [86] M. Gabriela M. Gomes, Marc Lipsitch, Andrew R. Wargo, Gael Kurath, Carlota Rebelo, Graham F. Medley, and Antonio Coutinho. A missing dimension in measures of vaccination impacts. *PLOS Pathogens*, 10(3):e1003849, 2014.
- [87] Kate E. Langwig, Andrew R. Wargo, Darbi R. Jones, Jessie R. Viss, Barbara J. Rutan, Nicholas A. Egan, Pedro Sá-Guimarães, Min Sun Kim, Gael Kurath, M. Gabriela M. Gomes, and Marc Lipsitch. Vaccine effects on heterogeneity in susceptibility and implications for population health management. *mBio*, 8(6):e00796–17, 2017.
- [88] Greg Dwyer, Joseph S. Elkinton, and John P. Buonaccorsi. Host heterogeneity in susceptibility and disease dynamics: Tests of a mathematical model. *The American Naturalist*, 150(6):685–707, 1997.
- [89] Warren A. Furumoto and Ray Mickey. A mathematical model for the infectivity-dilution curve of tobacco mosaic virus: Theoretical consideration. *Virology*, 32(2):216–223, 1967.
- [90] Frida Ben-Ami, Dieter Ebert, and Roland R. Regoes. Pathogen dose infectivity curves as a method to analyze the distribution of host susceptibility: a quantitative assessment of maternal effects after food stress and pathogen exposure. *The American Naturalist*, 175(1):106–115, 2010.

- [91] Jessica G. King, Caetano Souto-Maior, Larissa M. Sartori, Rafael Maciel-de-Freitas, and M. Gabriela M. Gomes. Variation in Wolbachia effects on Aedes mosquitoes as a determinant of invasiveness and vectorial capacity. *Nature Communications*, 9(1):1483, 2018.
- [92] Delphine Pessoa, Caetano Souto-Maior, Erida Gjini, João S. Lopes, Bruno Ceña, Cláudia T. Codeço, and M. Gabriela M. Gomes. Unveiling time in dose-response models to infer host susceptibility to pathogens. *PLOS Computational Biology*, 10(8):e1003773, 2014.
- [93] Fuminari Miura, Don Klinkenberg, and Jacco Wallinga. Quantifying the individual variation in susceptibility to endemic coronavirus and SARS-CoV-2 with human challenge trials. *Epidemiology*, 35(1):113–117, January 2024.
- [94] C. A. McGilchrist and C. W. Aisbett. Regression with frailty in survival analysis. *Biometrics*, 47(2):461–466, 1991.
- [95] Philip Hougaard. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, 71(1):75–83, 1984.
- [96] Alessandro Margheri, Carlota Rebelo, and M. Gabriela M. Gomes. On the correlation between variance in individual susceptibilities and infection prevalence in populations. *Journal of Mathematical Biology*, 71(6):1643–1661, 2015.
- [97] Tamás Tekeli, Andrea Pugliese, and Cinzia Soresina. SEIR models with host heterogeneity: theoretical aspects and applications to seasonal influenza dynamics. *arXiv preprint arXiv:2511.18142*, 2025.
- [98] Philip Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.
- [99] David D. Hanagal. Positive stable shared frailty models based on additive hazards. *Statistics in Biosciences*, 13(3):431–453, 2021.
- [100] Blavatnik School of Government, University of Oxford. Oxford COVID-19 Government Response Tracker (OxCGRT). <https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker>. Accessed 19 February 2026.
- [101] Stan Development Team. *Stan User’s Guide*, 2025. Version 2.37.
- [102] Andrew Gelman. Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, 48(3):432–435, 2006.

- [103] Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534, 2006.
- [104] Cora J. M. Maas and Joop J. Hox. Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3):86–92, 2005.
- [105] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013.
- [106] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, 2017.