

# SEQUENTIAL MONTE CARLO IN FINANCIAL ECONOMETRICS: FROM INFERENCE TO COMBINATORIAL OPTIMIZATION

A dissertation submitted in partial fulfillment of the requirements for the degree of

**PHD IN BUSINESS ADMINISTRATION**

and for the degree of

**DOCTEUR EN SCIENCES DE GESTION**

**DE L'ECOLE DOCTORALE**

«ECONOMIE, MANAGEMENT, MATHÉMATIQUES, PHYSIQUE ET SCIENCES INFORMATIQUES »

ED 405

CY CERGY PARIS UNIVERSITÉ

Presented and defended publicly on the 5$^{th}$ of June 2026 by

**Patrick TANG**

**JURY**

| | | |
|---|---|---|
| **Andras Fulop** | Supervisor | Professor, ESSEC Business School |
| **Jeremy Heng** | Supervisor | Associate Professor, ESSEC Business School |
| **Nicolas Chopin** | Referee | Professor, ENSAE |
| **Anna Simoni** | Referee | Professor, ENSAE |
| **Veronika Czellar** | President | Professor, SKEMA Business School |
| **Elise Gourier** | Examiner | Associate Professor, ESSEC Business School |

Copyright © 2026 Patrick Tang
All rights reserved.

# Abstract

This thesis studies the flexibility and practical relevance of Sequential Monte Carlo (SMC) methods as a unified computational framework for financial econometrics. It shows how a single inferential paradigm can address two fundamentally different classes of problems: parameter inference in nonlinear dynamic systems and high-dimensional combinatorial optimization.

The first contribution develops an SMC-based strategy for Bayesian estimation of long-run risk asset pricing models with mixed-frequency data. Standard pseudo-marginal methods become computationally infeasible in high-signal environments due to unstable likelihood estimates. The proposed approach replaces the unbiased likelihood requirement with a smooth approximation based on Gaussian filtering and common random numbers, producing a stable likelihood surface and drastically reducing computational cost. Empirical results on U.S. macro-financial data indicate that the COVID-19 period is best interpreted as a temporary shock rather than a persistent shift in preference or volatility parameters.

The second contribution reformulates sparse predictor selection in credit recovery modelling as a sampling problem over a discrete space. Using tempered SMC, the algorithm efficiently explores multimodal model landscapes and enables a consistent comparison between competing model specifications. The results show that incorporating firm-specific variables substantially improves predictive performance, and that models favored by likelihood criteria do not necessarily yield superior predictive accuracy or calibration.

Overall, the thesis demonstrates that SMC provides a general computational backbone across heterogeneous econometric tasks, but its effectiveness depends critically on tailoring the particle system and proposal mechanisms to the structure of each economic problem.

# Résumé

Cette thèse étudie la flexibilité théorique et l'utilité pratique des méthodes de Sequential Monte Carlo (SMC) comme cadre de calcul unifié pour l'économétrie financière moderne. Elle montre qu'un même paradigme d'inférence peut être utilisé pour traiter deux classes de problèmes fondamentalement différentes : l'estimation de paramètres dans des systèmes dynamiques non linéaires et l'optimisation combinatoire de grande dimension.

La première contribution propose une stratégie d'estimation bayésienne pour les modèles d'asset pricing de type long-run risk utilisant des données observées à fréquences mixtes, telles que la consommation trimestrielle et les rendements mensuels. Dans ces environnements riches en information, les méthodes pseudo-marginales classiques deviennent rapidement impraticables en raison de l'instabilité des estimations de vraisemblance. L'approche développée remplace l'exigence d'une vraisemblance non biaisée par une approximation lisse fondée sur un filtrage gaussien et l'utilisation de nombres aléatoires communs, ce qui stabilise la surface de vraisemblance et réduit fortement le coût de calcul. L'application aux données macro-financières américaines suggère que la période liée au COVID-19 correspond davantage à un choc transitoire qu'à une modification permanente des préférences ou de la dynamique de volatilité.

La seconde contribution reformule la sélection parcimonieuse de variables explicatives dans la modélisation des taux de recouvrement de crédit comme un problème d'échantillonnage dans un espace discret complexe. Grâce à un schéma SMC tempéré, l'algorithme explore efficacement des paysages de modèles multimodaux et permet une comparaison cohérente entre différentes spécifications concurrentes. Les résultats montrent que l'intégration d'informations spécifiques aux entreprises améliore significativement les performances prédictives et que les modèles favorisés par les critères de vraisemblance ne sont pas nécessairement les mieux calibrés ni les plus précis en prévision.

Dans l'ensemble, la thèse met en évidence que les méthodes SMC constituent une base computationnelle générale pour des tâches économétriques hétérogènes, mais que leur efficacité dépend fortement de l'adaptation du système de particules et des mécanismes de proposition à la structure économique du problème étudié.

# Acknowledgements

With the completion of this thesis, I would like to express my sincere gratitude to my doctoral advisors, Prof. Andras Fulop and Prof. Jeremy Heng, for their guidance and patience throughout my PhD. Working with you, including time spent in Singapore and Budapest, has been a defining part of this journey. Over the years, I learned not only technical aspects of research but also the importance of rigor, clarity, intellectual discipline, and ownership in pursuing ideas. I am equally grateful for the generous sharing of knowledge and expertise, as well as for our conversations and the perspective you shared on the importance of family and ways of life. These lessons will remain valuable to me in my professional life long after this dissertation.

I thank my co-authors, Prof. Junye Li and Prof. Jin-Chuan Duan, for their collaboration. Their responsiveness and insights from different perspectives strengthened this work, and I deeply appreciate the time and expertise they invested in it.

I would also like to acknowledge ESSEC Business School and the professors in the Department of Information Systems, Data Analytics and Operations. The coursework during the first two years provided a strong foundation and shaped how I approach problems and structure ideas.

I am sincerely grateful to the members of my jury, Prof. Nicolas Chopin, Prof. Elise Gourier, Prof. Anna Simoni, and Prof. Veronika Czellar, for their time, careful reading, and valuable feedback. I am particularly grateful to Prof. Nicolas Chopin, whose course provided specialized training and strengthened my understanding of Monte Carlo methods. I also thank Prof. Elise Gourier for attending multiple jury meetings at different stages of my PhD and for her insightful advice.

I am grateful to the PhD community around me for the conversations, shared experi-

ences, and mutual support that made this journey both intellectually rich and personally meaningful.

My appreciation also goes to Lina Prevost, Christine Gil, and Audrey Soulet from the ESSEC PhD Office, and Lalia Sy at CY Cergy Paris, whose assistance and steady support made the administrative process far smoother than it would otherwise have been.

To my mother and my grandfather, this PhD journey allowed me to grow quickly and to understand many things more clearly. Above all, it made me realize how much support and love had been present since the very beginning of my life. You gave me the foundation to stand on, and your encouragement, together with countless small gestures along the way, shaped who I am today.

Finally, to my wife. You joined me halfway through this journey, but you changed its destination. Your presence brought balance and joy at the moment it was most needed, and made finishing this work meaningful beyond the thesis itself.

This journey has taken me across Paris, Singapore, and Budapest. I am grateful for the people I met and the lessons I learned along the way. I leave this period with a deep sense of gratitude, and with excitement for what comes next.

# Contents

**Abstract** i

**Résumé** ii

**Acknowledgements** iv

**1 Introduction** 1

**2 Mathematical Foundations and Literature Review** 4

2.1 Background: MCMC, IS, and the Need for Sequential Methods . . . . . 4

2.2 The General SMC Framework . . . . . 5

2.2.1 Notation Conventions . . . . . 6

2.2.2 Algorithmic Structure and Weights . . . . . 6

2.2.3 The Path of Distributions . . . . . 7

2.2.4 Stability Mechanisms: Resampling, ESS, and Rejuvenation . . . . . 7

2.3 SMC for State-Space Filtering . . . . . 8

2.3.1 The Intractability Challenge and the Kalman Filter Exception . . . 9

2.3.2 Mapping SMC to the Filtering Problem . . . . . 10

2.3.3 Implementation Challenges and the Curse of Dimensionality . . . . 11

2.3.4 Bootstrap Particle Filter . . . . . 11

2.3.5 Auxiliary Particle Filter . . . 12
2.3.6 Unscented Particle Filter . . . 12
2.3.7 Rao-Blackwellised Particle Filter . . . 13
2.4 SMC for Parameter Inference . . . 13
2.4.1 The MCMC Approach: Particle Metropolis-Hastings . . . 14
2.4.2 The SMC Approach: SMC$^2$ . . . 14
2.4.3 Computational Challenges and Variance Reduction . . . 15
2.4.4 Applications in Macroeconomics and Finance . . . 16
2.5 SMC for Optimization and Selection . . . 17
2.5.1 The Tempered SMC Sampler . . . 17
2.5.2 The Rejuvenation Step: Boosting Particle Diversity . . . 18
2.5.3 Application to Discrete and Combinatorial Spaces . . . 19
2.6 Conclusion: Bridging Theory and Application . . . 19

**3 Bayesian Parameter Estimation in Long-Run Risk Models with Mixed-Frequency Data 21**

3.1 Introduction . . . 22
3.2 A Consumption-based Long-Run Risk Model . . . 26
3.2.1 Fundamentals . . . 27
3.3 Long-Run Risk Model as State-Space Model . . . 29
3.3.1 Re-write Long-Run Risk Model . . . 29
3.3.2 Model with Mixed-Frequency Data . . . 31
3.3.3 Extended state space . . . 32
3.4 Econometric Framework . . . 36
3.4.1 Pseudo-Marginal Methods and Their Limitations . . . 36

3.4.2 Using Common Random Numbers and Smoothing the Likelihood . 37
3.4.3 Refinements and Practical Implementation . . . . . 39
3.4.4 Empirical Validation of the Inference Methodology . . . . . 40
3.4.5 SMC Configuration . . . . . 46
3.4.6 Monte Carlo Simulation Studies . . . . . 47
3.5 Data and Results . . . . . 50
3.5.1 Data Description and Summary Statistics . . . . . 50
3.5.2 Filtered States . . . . . 52
3.5.3 Mixed-Frequency Estimation: COVID Impact and Structural Comparison . . . . . 54
3.6 Conclusion . . . . . 56

**4 Credit Recovery Modelling via Sequential Monte Carlo Optimization 58**
4.1 Introduction . . . . . 59
4.2 Recovery Modelling . . . . . 62
4.2.1 Beta Regression . . . . . 63
4.2.2 Logit-Transformed Linear Regression . . . . . 64
4.3 Variable Selection using SMC . . . . . 65
4.4 Empirical Application . . . . . 68
4.4.1 Data Preparation . . . . . 71
4.4.2 Predictive Performance . . . . . 74
4.4.3 Conditional Calibration and Distributional Diagnostics . . . . . 75
4.4.4 Economic Implications . . . . . 77
4.5 Conclusion . . . . . 80

**5 Conclusion and Future Directions 82**

5.1 Summary of Findings and Contributions . . . 82
5.2 Synthesis of Empirical Approaches and Challenges . . . 82
5.3 Key Empirical Results and Implications . . . 84
5.4 Limitations and Future Research . . . 85

**A Appendix to Chapters 2** **94**

**B Appendix to Chapter 3** **101**

B.1 Derivation of the Autoregressive-gamma Transition . . . 101
B.1.1 Proposal distributions . . . 103
B.2 Observation Density of the Balanced-Frequency Model . . . 104
B.2.1 Joint Distribution . . . 105
B.2.2 Measurement Density . . . 105
B.3 Sequential Monte Carlo for State Filtering and Likelihood Estimation. . . . 107
B.4 Pseudo Code of S-UPF-CRN-SMC . . . 108
B.4.1 UKF-Based Proposal for Each Particle . . . 108
B.4.2 S-UPF-CRN: Smoothed Unscented Particle Filter with CRNs . . . 108
B.4.3 S-UPF-CRN-SMC: SMC Sampler over $\theta$ Using S-UPF-CRN . . . 109
B.5 Empirical Validation with Covid-19 Data . . . 110
B.6 Additional Log-Likelihood Profiles . . . 112
B.7 Additional Monte Carlo Simulation Results . . . 112

**C Appendix to Chapter 4** **114**

C.1 Density-tempered SMC . . . 114
C.1.1 Initialization . . . 114
C.1.2 Reweighting and Resampling . . . 115

C.1.3 Support Boosting . . . . . 115
C.1.4 Adding or Trimming Regressors . . . . . 117
C.1.5 Determining the Right Number of Regressors . . . . . 118
C.2 Results of Unfiltered Data . . . . . 119
C.3 Out-of-Sample Log-Likelihood Comparisons . . . . . 121

# Chapter 1

# Introduction

State estimation in economics and finance has historically relied on the Kalman Filter. While optimal for linear and Gaussian systems, this methodology proves inadequate for modern financial models characterized by nonlinear dynamics and non-Gaussian shocks. To address these limitations, Kim et al. (1998) introduced particle filters to the economics literature. Since then, their use has expanded from estimating stochastic volatility to solving complex Dynamic Stochastic General Equilibrium (DSGE) models (Fernández-Villaverde and Rubio-Ramírez, 2005; Fernández-Villaverde and Rubio-Ramírez, 2007; Creal, 2012).

However, particle filtering is merely a specific instance of a broader class of algorithms: Sequential Monte Carlo (SMC). Fundamentally, SMC is a framework for sampling from a sequence of probability distributions. This sequence need not represent physical time; it can represent an artificial evolution of densities, a property that allows SMC to solve static, non-sequential problems as well (Del Moral, 2004; Chopin and Papaspiliopoulos, 2020).

This thesis leverages the theoretical flexibility of the SMC framework to introduce novel algorithmic solutions that transcend the limitations of standard econometric tools. First, we aim to make Bayesian inference computationally feasible in complex, mixed-frequency environments where traditional methods fail. Second, we investigate whether a probabilistic approach to selecting variables can reveal structural trade-offs in credit risk modelling that are missed by simpler algorithms. In Chapter 2, we provide a unified exposition of this framework, demonstrating how the underlying Feynman-Kac models can be adapted to solve problems ranging from continuous latent state estimation to discrete variable selection. Specifically, we distinguish between two methodological paths: first, the path of partial

posteriors (physical time), used for state-space filtering in high-signal time-series models; and second, the geometric path (artificial tempering), used for static parameter inference and combinatorial optimization.

## Contribution 1: Inference in Mixed-Frequency Long-Run Risk Models

The first application (Chapter 3) addresses Bayesian parameter estimation in long-run risk (LRR) models. These models are characterized by nonlinear dynamics and are estimated here using mixed-frequency data, combining quarterly consumption and dividend growth with monthly returns and interest rates.

A major computational bottleneck in this literature is that observations are highly informative relative to the state transition. In this high-signal regime, standard pseudo-marginal methods—such as Particle MCMC (PMCMC) and SMC$^2$—require unbiased likelihood estimates, which leads to prohibitive variance. Achieving the stability required for standard MCMC convergence would necessitate runtimes exceeding 200 days.

We propose and implement an alternative estimation strategy: S-UPF-CRN-SMC. This method replaces the strict pseudo-marginal requirement with a smooth likelihood approximation based on Gaussian filtering and common random numbers (CRNs). By inducing strong correlations across likelihood evaluations, this approach stabilizes posterior inference and reduces runtime by several orders of magnitude.

We apply this framework to U.S. data spanning 1959 to 2022. The use of mixed-frequency data allows for the identification of short-run risk components that are often obscured in temporally aggregated samples. Our results indicate that including the COVID-19 period leads to marked reductions in estimated risk aversion and volatility persistence. These findings suggest that the pandemic represented a transitory shock rather than a permanent structural shift, highlighting the empirical value of high-frequency data in identifying temporary shifts in macro-financial risk.

## Contribution 2: Combinatorial Optimization for Credit Recovery

The second application (Chapter 4) adapts the SMC framework for static optimization. We address the problem of feature selection in credit recovery rate modelling, framing it as an $\ell_0$-penalized optimization problem where a parsimonious set of predictive features must be selected from a large potential set.

Standard stepwise regression methods often become trapped in local optima when the predictor space is high-dimensional and collinear. We reformulate variable selection as a sampling problem from a complex discrete distribution over predictor combinations. We solve this using tempered SMC, where the sequence of target distributions is defined by an increasing temperature parameter. This allows the sampler to traverse the multimodal landscape of potential models effectively.

We exploit the flexibility of this optimization engine to conduct a consistent evaluation of two competing specifications—the logit-transformed heteroskedastic model and the beta model—applied to a dataset of U.S. dollar-denominated bonds. The analysis reveals a systematic discrepancy between evaluation metrics: while likelihood-based metrics favor the logit specification, the beta model yields superior predictive accuracy (higher $R^2$) and passes formal calibration tests. Furthermore, we demonstrate that incorporating firm-specific variables significantly enhances model performance compared to macro-only baselines. Ultimately, the analysis confirms that the SMC framework can deliver robust, interpretable models even when applied to large, collinear predictor sets.

# Chapter 2

# Mathematical Foundations and Literature Review

## 2.1 Background: MCMC, IS, and the Need for Sequential Methods

Approximating probability distributions has historically relied on two main Monte Carlo paradigms: Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS).

The fundamental objective of MCMC is to construct a Markov chain with a stationary distribution matching the target $\pi(x)$. While widely adopted through algorithms like the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970; Diaconis, 2009) (Algorithm 1), significant advances have been made to improve exploration. Notably, gradient-based methods such as the Metropolis-Adjusted Langevin Algorithm (MALA) and Hamiltonian Monte Carlo (HMC) exploit the derivative of the target density to guide the chain toward high-probability regions, reducing random walk behaviour (Roberts and Tweedie, 1996; Neal, 2011) (Algorithm 2, 3).

However, despite these improvements, standard MCMC methods face fundamental algorithmic limitations, particularly in the context of complex financial models. Structurally, their iterative nature precludes straightforward parallelization, creating a computational bottleneck. Furthermore, in terms of exploration, a single chain may become trapped in local

modes within multimodal distributions. Additionally, practical implementation requires a burn-in period to reach stationarity—a state where convergence diagnosis is non-trivial—and performance remains highly sensitive to the tuning of proposal distributions.

Moreover, estimating the normalizing constant $Z$, a critical component for Bayesian model comparison, remains computationally challenging within the MCMC framework (Dai et al., 2022). Conversely, Importance Sampling (IS) facilitates parallelization and yields a direct estimate of $Z$. Nevertheless, standard IS becomes ineffective in high-dimensional settings due to the inevitable variance explosion of importance weights (Agapiou et al., 2017).

By the late 1990s, research focused on unifying these paradigms. SMC methods, originally developed as particle filters for state-space models (Gordon et al., 1993), were generalized to static problems under the framework of SMC Samplers (Chopin, 2002; Del Moral et al., 2006). This unified framework synthesizes the strengths of its predecessors: it operates on a population of particles (like IS), enabling extensive parallelization, and it employs Markov kernels for exploration (like MCMC) but avoids the burn-in period. Furthermore, the particle population makes it reasonably easy to develop adaptive samplers, allowing tuning parameters to be calibrated automatically using the current sample. This foundational structure enables the sequential approximation of complex distributions, which is detailed in the general framework below.

## 2.2 The General SMC Framework

SMC methods approximate a sequence of probability distributions $\{\pi_t\}_{t=0}^{T}$ defined on a common measurable space $(\mathsf{X}, \mathcal{X})$. Each $\pi_t$ is specified via an unnormalized density $\gamma_t$ with respect to a base measure $\mu$:

$$\pi_t(\mathrm{d}x) = \frac{\gamma_t(x)}{Z_t}\,\mu(\mathrm{d}x), \qquad Z_t := \int_{\mathsf{X}} \gamma_t(x)\,\mu(\mathrm{d}x), \quad \text{for } t = 0, \ldots, T. \tag{2.1}$$

The normalizing constants $Z_t$ are typically unknown in practice. The Feynman–Kac formulation provides a unifying theoretical framework for such sequences and underpins convergence results, central limit theorems, and variance bounds for SMC estimators (Del Moral, 2004).

### 2.2.1 Notation Conventions

Throughout this chapter, the interpretation of indices and parameters adapts to the specific problem context to align with the standard conventions of the respective fields. Specifically, the index $t$ assumes distinct roles depending on the algorithmic goal: in filtering and parameter inference (Sections 2.3 and 2.4), it denotes physical time, whereas in optimization (Section 2.5), it represents an algorithmic iteration or temperature level.

Similarly, while the generic SMC backbone maintains consistent notation, the specific application chapters adopt symbols standard to their domains to facilitate readability. Certain symbols may carry context-specific meanings. For instance, the symbol $\phi$ is used in Chapter 3 to denote volatility or persistence parameters within the state transition dynamics, yet re-emerges in Chapter 4 to denote the precision parameter of the Beta distribution. Therefore, readers are advised to interpret these symbols within the local context of each chapter.

### 2.2.2 Algorithmic Structure and Weights

SMC algorithms maintain a weighted particle approximation of each $\pi_t$. At $t = 0$, the algorithm samples particles from $\pi_0$ (or a tractable approximation) and constructs the empirical measure $\pi_0^N(\mathrm{d}x) := \sum_{i=1}^N w_0^{(i)}\, \delta_{x_0^{(i)}}(\mathrm{d}x)$, with normalized weights $w_0^{(i)} = 1/N$. For $t \geq 1$, the approximation $\pi_{t-1}^N$ is transformed into $\pi_t^N$ through three steps:

1. Propagation: Each particle $x_{t-1}^{(i)}$ is propagated through a forward Markov kernel $M_t(x_{t-1}, \cdot)$ with density $m_t(x_t \mid x_{t-1})$ to propose a new state $x_t^{(i)}$.

2. Weighting: Proposed particles are reweighted to correct for the discrepancy between the proposal and the new target $\pi_t$.

3. Resampling (rejuvenation): If the weight distribution is too concentrated, a resampling step discards low-weight particles and replicates high-weight particles.

The incremental importance weight on the joint space $(x_{t-1}, x_t)$ is derived by introducing a backward kernel $L_{t-1}(x_t, \cdot)$ with density $\ell_{t-1}(x_{t-1} \mid x_t)$ as (Del Moral et al., 2006):

$$w_t(x_{t-1}, x_t) := \frac{\gamma_t(x_t)\, \ell_{t-1}(x_{t-1} \mid x_t)}{\gamma_{t-1}(x_{t-1})\, m_t(x_t \mid x_{t-1})}. \tag{2.2}$$

The estimator of the ratio of normalizing constants, $\widehat{Z_T/Z_0}$, is formed from the product of the average unnormalized importance weights:

$$\widehat{Z_T/Z_0} := \prod_{t=1}^{T} \left( \frac{1}{N} \sum_{i=1}^{N} w_t\big(x_{t-1}^{(a_i)}, x_t^{(i)}\big) \right), \tag{2.3}$$

where $a_i$ denotes the ancestor index after resampling. When $Z_0 = 1$, this estimator $\widehat{Z}_T$ is unbiased and is key for Bayesian model comparison and evidence estimation.

### 2.2.3 The Path of Distributions

The sequence of intermediate unnormalized densities $\{\gamma_t\}$ is central to bridging a tractable initial distribution $\pi_0$ and the final target $\pi_T$. The construction of this sequence depends fundamentally on the problem structure, giving rise to two principal paths:

1. Path of partial posteriors: The sequence arises naturally as each subsequent target incorporates one additional observation. The target at step $t$ is $\pi_t(x) = p(x \mid y_{1:t})$. This is the foundation of particle filters (Section 2.3).

2. Geometric path (tempering): The sequence is defined artificially to gradually introduce the complexity of the final target. Intermediate densities take the form

$$\gamma_t(x) = \gamma_0(x)^{1-\lambda_t}\, \gamma_T(x)^{\lambda_t}, \qquad 0 = \lambda_0 < \cdots < \lambda_T = 1,$$

   where $\lambda_t$ are temperature parameters. This path is used in SMC Samplers for parameter inference and optimization (Sections 2.4 and 2.5).

### 2.2.4 Stability Mechanisms: Resampling, ESS, and Rejuvenation

Repeated application of importance weighting inevitably leads to weight degeneracy, where a small number of particles carries most of the weight. SMC samplers employ two main mechanisms to mitigate this phenomenon.

First, degeneracy is monitored using the effective sample size (ESS),

$$\mathrm{ESS}_t := \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2}, \qquad w_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{j=1}^{N} \omega_t^{(j)}}, \tag{2.4}$$

where $\omega_t^{(i)}$ are the unnormalized weights. When $\mathrm{ESS}_t$ falls below a threshold (often $N/2$), a resampling step is triggered to discard low-weight particles and replicate high-weight particles. A small ratio $\mathrm{ESS}_t/N$ signals a pronounced mismatch between the proposal and the target.

Furthermore, this diagnostic is crucial for tuning the particle count $N$. In pseudo-marginal MCMC (Section 2.4), persistent ESS collapse indicates that improving the proposal kernels $M_t$ is essential to maintain a stable variance of the log-likelihood estimator.

However, resampling induces sample impoverishment, where identical copies of high-weight particles proliferate. To counter this, SMC samplers often include a rejuvenation or move step that applies an MCMC kernel $K_t$ (e.g., MH) invariant with respect to $\pi_t$ to each resampled particle. This step jitters cloned particles while preserving the target distribution.

## 2.3 SMC for State-Space Filtering

We now specialize the generic SMC framework (Section 2.2) to the time-series context, where the SMC index $t$ becomes physical time, and the path of distributions arises naturally as a sequence of partial posteriors (Section 2.2.3). This specialization is known as the particle filter (PF), the algorithm for filtering latent states in state-space Models.

State-space models (SSMs), or hidden Markov models, are a prominent class of time series models with applications spanning ecology, econometrics, engineering, and neuroscience (Doucet et al., 2001; Cappé et al., 2005). Examples include stochastic volatility models where $X_t$ represents asset volatility and $Y_t$ log-returns (Kim et al., 1998); biochemical network models where $X_t$ denotes species populations (Wilkinson, 2012); and neuroscience models where $X_t$ determines a neuron's stimulus–response function (Paninski et al., 2010).

Formally, a SSM comprises an unobserved $\mathsf{X}$-valued latent Markov process $\{X_t\}_{t\geq 0}$ and an observed $\mathsf{Y}$-valued process $\{Y_t\}_{t\geq 0}$. We consider a nonlinear, non-Gaussian SSM defined

by densities w.r.t. base measure $\mu$:

$$X_t \sim f_\theta(x_t|x_{t-1}), \qquad Y_t \sim g_\theta(y_t|x_t), \qquad t = 1{:}T, \tag{2.5}$$

with initial density $X_0 \sim \mu_\theta(x_0)$. The objective is inferring latent states $X_t$ given observations $y_{1:T}$. Key tasks include:

- Filtering: Computing the filtering density $p_\theta(x_t|y_{1:t})$.
- State Prediction: Computing the predictive density $p_\theta(x_{t+h}|y_{1:t})$ for $h \geq 1$.
- Smoothing: Computing the smoothing density $p_\theta(x_{0:t}|y_{1:T})$.

Analytically, the filtering density satisfies the prediction and update recursions:

$$\text{Predict:} \quad p_\theta(x_t|y_{1:t-1}) = \int_{\mathsf{X}} f_\theta(x_t|x_{t-1})\, p_\theta(x_{t-1}|y_{1:t-1})\, \mu(\mathrm{d}x_{t-1}), \tag{2.6}$$

$$\text{Update:} \quad p_\theta(x_t|y_{1:t}) = \frac{g_\theta(y_t|x_t)\, p_\theta(x_t|y_{1:t-1})}{p_\theta(y_t|y_{1:t-1})}, \tag{2.7}$$

where $p_\theta(y_t|y_{1:t-1}) = \int_{\mathsf{X}} g_\theta(y_t|x_t)\, p_\theta(x_t|y_{1:t-1})\, \mu(\mathrm{d}x_t)$ is the predictive likelihood.

### 2.3.1 The Intractability Challenge and the Kalman Filter Exception

While this two-step recursion offers an optimal theoretical solution, it is analytically intractable for most practical applications. The bottleneck lies in the prediction integral (2.6). Unless $p(x_{t-1}|y_{1:t-1})$ and $p(x_t|x_{t-1})$ belong to specific conjugate families, no closed-form solution exists.

This issue is cumulative. Even if $p(x_0)$ is simple (e.g., Gaussian), $p(x_1|y_1)$ typically becomes a complex, multimodal distribution after one step. This intractability historically compelled practitioners to rely on approximations such as the Extended Kalman Filter (EKF).

A notable exception is the linear Gaussian state-space model, defined as:

- State Equation: $x_t = \mathbf{A}_t x_{t-1} + v_t$, where $v_t \sim \mathcal{N}(0, \mathbf{Q}_t)$

- Observation Equation: $y_t = \mathbf{C}_t x_t + e_t$, where $e_t \sim \mathcal{N}(0, \mathbf{R}_t)$

In this specification, $\mathbf{A}_t$ governs the deterministic evolution of the latent state, while $\mathbf{C}_t$ maps this state to the observed data. The terms $v_t$ and $e_t$ represent independent Gaussian noise processes, specifically the structural shock and measurement error, characterized by covariance matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$, respectively. Here, the Gaussian distribution is closed under prediction and update operations. Thus, if $p(x_0)$ is Gaussian, $p(x_t|y_{1:t})$ remains Gaussian for all $t$. The optimal Bayesian filter simplifies to the Kalman Filter (Welch and Bishop, 1995), which recursively updates the mean and covariance.

### 2.3.2 Mapping SMC to the Filtering Problem

For general nonlinear, non-Gaussian cases, SMC provides a numerical approximation. Time $t$ acts as the SMC index, targeting joint smoothing densities

$$\pi_t(\mathrm{d}x_{0:t}) = p_\theta(x_{0:t}|y_{1:t})\, \mu(\mathrm{d}x_{0:t}).$$

The forward kernel $M_t$ is the proposal density $q_t(x_t|x_{t-1}, y_t)$, and a backward kernel $\ell_{t-1}(x_{t-1}|x_t)$ is implicitly defined. Substituting the target $\gamma_t(x_{0:t}) = p_\theta(x_{0:t}, y_{1:t})$ into the generic weight (2.2) yields the standard particle filter incremental weight:

$$w_t(x_{t-1}, x_t) \propto \frac{g_\theta(y_t|x_t)\, f_\theta(x_t|x_{t-1})}{q_t(x_t|x_{t-1}, y_t)}. \tag{2.8}$$

The minimum variance proposal is the true state conditional $q_t^*(x_t|x_{t-1}, y_t) = p_\theta(x_t|x_{t-1}, y_t) \propto g_\theta(y_t|x_t)\, f_\theta(x_t|x_{t-1})$. Using this proposal minimizes weight dependency on $x_t$, reducing degeneracy (Doucet et al., 2001).

In the linear Gaussian state-space context, a particle filter using the optimal proposal $q_t^*$ and optimal resampling recovers the KF mean estimate (Pitt and Shephard, 1999). This motivates designing proposals $q_t$ to approximate $p_\theta(x_t|x_{t-1}, y_t)$. The unbiased likelihood estimator $\widehat{Z}_T = \widehat{p}_\theta(y_{1:T})$ is derived via (2.3) (see Section 2.4). The optimal proposal $q_t^*$ is the theoretical ideal but is typically as intractable as the filtering problem itself. Consequently, standard particle filter algorithms are essentially practical approximations of this optimal proposal.

### 2.3.3 Implementation Challenges and the Curse of Dimensionality

The practical performance of particle filters depends critically on the quality of the proposal distribution $q_t$ relative to the true target. Two primary phenomena compromise this stability: the dimensionality of the state-space and the informativeness of observations. These two theoretical hurdles constitute the primary computational challenges addressed in this thesis: the highly informative observations encountered in the LRR application (Chapter 3) and the high-dimensional search space inherent to the credit recovery application (Chapter 4).

In high-dimensional state-spaces ($d_x \gg 10$), the volume of the space increases exponentially. The region where the proposal density $q_t$ and the likelihood $g_\theta$ significantly overlap becomes infinitesimally small relative to the total volume (Chopin and Papaspiliopoulos, 2020). Unless the number of particles $N$ increases exponentially with $d_x$, the vast majority of particles will land in low-probability regions, leading to immediate weight collapse where a single particle dominates the ensemble (Kantas et al., 2015).

A similar collapse occurs even in low dimensions if the observations are highly informative (i.e., the likelihood $g_\theta(y_t|x_t)$ is extremely peaked compared to the state transition $f_\theta(x_t|x_{t-1})$). This is a frequent challenge in financial applications with high signal-to-noise ratios. A simple proposal like the bootstrap particle filter, which propagates states based solely on dynamics, will consistently miss the narrow high-likelihood region, resulting in a negligible ESS.

To address these challenges, advanced variants have been developed to enhance sampling efficiency. Guided proposals, such as the auxiliary and unscented particle filters, incorporate the current observation $y_t$ into the proposal mechanism; these filters concentrate particles into the high-likelihood regions, maintaining diversity without requiring an exponential increase in $N$. Complementarily, where the model structure permits, dimension reduction techniques like the Rao-Blackwellised particle filter exploit analytical marginalization to explicitly lower the dimensionality of the state-space requiring Monte Carlo approximation.

### 2.3.4 Bootstrap Particle Filter

The bootstrap particle filter (BPF) uses the state transition as the proposal (Algorithm 4):

$$q_t(x_t|x_{t-1}, y_t) = f_\theta(x_t|x_{t-1})$$

. The weight (2.8) simplifies to the observation density:

$$w_t(x_{t-1}, x_t) \propto g_\theta(y_t|x_t). \tag{2.9}$$

The BPF is effective when transitions $f_\theta$ are informative or observations $g_\theta$ are diffuse, offering minimal computational overhead (Gordon et al., 1993). However, if $g_\theta$ is sharply peaked (informative observations) relative to $f_\theta$, particles may land in low-likelihood regions, causing weight collapse. Regarding implementation, adaptive low-variance resampling is typically used (e.g., when ESS $< N/2$).

### 2.3.5 Auxiliary Particle Filter

The auxiliary particle filter (APF) reduces degeneracy by incorporating $y_t$ prior to propagation (Pitt and Shephard, 1999) , as detailed in Algorithm 5. It uses an approximation $m_t(x_{t-1}) \approx p_\theta(y_t|x_{t-1})$ to pre-select ancestors likely to yield high-likelihood states. The final weight corrects for the proposal $q_t$ and auxiliary weight $m_t$:

$$w_t(x_{t-1}^{(a_i)}, x_t^{(i)}) \propto \frac{g_\theta(y_t|x_t^{(i)})\, f_\theta(x_t^{(i)}|x_{t-1}^{(a_i)})}{q_t(x_t^{(i)}|x_{t-1}^{(a_i)}, y_t)\, m_t(x_{t-1}^{(a_i)})}. \tag{2.10}$$

This approach is valuable when observations are highly informative and a cheap approximation $m_t$ is available. Conversely, a poor $m_t$ can misguide particles, potentially increasing variance.

### 2.3.6 Unscented Particle Filter

The unscented particle filter (UPF) employs an unscented Kalman filter (UKF) per particle to generate a Gaussian proposal $q_t = \mathcal{N}(\mu_{t|t}^{(i)}, \Sigma_{t|t}^{(i)})$ approximating the local filtering density, proposed by Van der Merwe et al. (2000) (see Algorithm 6). The UKF propagates sigma points through $f_\theta$ and $g_\theta$ to estimate the proposal moments conditional on $y_t$. The weight is:

$$w_t(x_{t-1}^{(i)}, x_t^{(i)}) \propto \frac{g_\theta(y_t|x_t^{(i)})\, f_\theta(x_t^{(i)}|x_{t-1}^{(i)})}{\mathcal{N}\big(x_t^{(i)}; \mu_{t|t}^{(i)}, \Sigma_{t|t}^{(i)}\big)}. \tag{2.11}$$

The UPF reduces variance in nonlinear models where local Gaussian approximations hold, making it effective for sharp likelihoods and stable likelihood estimation in SMC$^2$. However, it suffers from poor performance if the true conditional density is strongly non-Gaussian and incurs a higher computational cost per particle.

### 2.3.7 Rao-Blackwellised Particle Filter

The Rao-Blackwellised particle filter (RBPF) applies when the state $x_t = (x_{t,l}, x_{t,nl})$ allows analytical marginalization of the linear state components $x_{t,l}$ (typically via a Kalman filter) conditional on the nonlinear state components $x_{0:t,nl}$ (Murphy and Russell, 2001; Doucet et al., 2001), as detailed in Algorithm 7.

It operates only on the nonlinear $x_{t,nl}$, while each particle carries sufficient statistics for the conditional Gaussian of the linear $x_{t,l}$. Analytically marginalizing the linear substructure reduces the sampling dimension, thereby lowering variance (Grisetti et al., 2007). Nevertheless, this method is restricted to conditionally linear-Gaussian structures and requires higher per-particle computation.

## 2.4 SMC for Parameter Inference

Often, the objective extends to estimating the static parameter vector $\theta$, where Bayesian inference assigns a prior $p(\theta)$ and targets the posterior distribution $p(\theta|y_{1:T})$. As detailed in the comprehensive review by Kantas et al. (2015), two primary strategies exist for sampling from this posterior. The first, particle Gibbs, samples from the joint posterior $p(\theta, x_{0:T}|y_{1:T})$ by alternating between conditional updates of the states and parameters. Alternatively, pseudo-marginal approaches marginalize out the latent states to target $p(\theta|y_{1:T})$ directly.

This thesis focuses on the latter. Ideally, one would employ a standard Metropolis-Hastings (MH) sampler to target $p(\theta|y_{1:T})$. However, this Ideal MH is impossible because evaluating the marginal likelihood $p_\theta(y_{1:T})$ requires integrating over all possible latent state paths—an intractable operation for nonlinear, non-Gaussian SSMs.

A conceptually simple alternative would be to augment the state vector within an SMC framework, defining $X'_t = (x_t, \theta)$. However, because fixed parameters lack process

noise ($\theta_t = \theta_{t-1}$), this approach fails systemically due to particle impoverishment: repeated resampling steps deplete the diversity of $\theta$ values, causing the ensemble to collapse to a single point (Kantas et al., 2015).

Consequently, successful inference requires algorithms that strictly separate state estimation from parameter exploration. This necessitates the pseudo-marginal framework. As formalized by Andrieu and Roberts (2009), we replace the intractable likelihood in the MH acceptance ratio with a non-negative unbiased estimator $\widehat{p}_\theta(y_{1:T})$ generated by the particle filter:

$$\widehat{p}_\theta(y_{1:T}) = \widehat{Z}_T \tag{2.12}$$

Although this estimator is noisy, the resulting algorithm targets a distribution on an extended space (including the particle auxiliary variables) that admits the exact posterior $p(\theta|y_{1:T})$ as its marginal. This effectively allows for exact Bayesian inference despite the use of approximate likelihoods.

### 2.4.1 The MCMC Approach: Particle Metropolis-Hastings

The most direct application of the pseudo-marginal principle is Particle Metropolis-Hastings (PMMH) (Andrieu et al., 2010) (Algorithm 8). This algorithm constructs a Markov chain for $\theta$. Given the current state $\theta^{(i-1)}$, it proposes a new candidate $\theta^*$, runs a particle filter to compute $\widehat{p}_{\theta^*}(y_{1:T})$, and accepts the move with probability:

$$\alpha(\theta^{(i-1)}, \theta^*) := \min\left\{1, \frac{\widehat{p}_{\theta^*}(y_{1:T})\, p(\theta^*)\, q(\theta^{(i-1)}|\theta^*)}{\widehat{p}_{\theta^{(i-1)}}(y_{1:T})\, p(\theta^{(i-1)})\, q(\theta^*|\theta^{(i-1)})}\right\}. \tag{2.13}$$

The resulting chain targets the joint distribution of $(\theta, \widehat{p}_\theta)$, ensuring the correct marginal for $\theta$. However, like all MCMC methods, PMMH is sequential and can be difficult to parallelize. Its efficiency is strictly governed by the variance of the log-likelihood estimator .

### 2.4.2 The SMC Approach: SMC$^2$

SMC$^2$ embeds particle filters within an outer SMC sampler (Chopin et al., 2013; Fulop and Li, 2013). Instead of a single chain, SMC$^2$ evolves a population of $\theta$-particles. Since the inner filters for each $\theta$-particle are independent, this approach is highly parallelizable.

Broadly, two main implementation strategies exist to manage this population.

The first variant, data tempering (Algorithm 9) targets the sequence of partial posteriors $p(\theta|y_{1:t})$ as data arrives. $N_\theta$ parameter particles manage their own inner particle filters to estimate incremental likelihoods $\widehat{p}(y_t|y_{1:t-1}, \theta_{t-1}^{(j)})$, which determine the outer weights:

$$\omega_t^{(j)} \propto W_{t-1}^{(j)} \cdot \widehat{p}(y_t|y_{1:t-1}, \theta_{t-1}^{(j)}). \tag{2.14}$$

When the outer ESS drops, the $\theta$-particles are rejuvenated using a PMMH kernel.

Alternatively, the density tempering approach ( Algorithm (10) bridges the prior $p(\theta)$ and the full posterior $p(\theta|y_{1:T})$ via a geometric path of tempered targets (Duan and Fulop, 2015):

$$\pi_k(\theta) \propto p(\theta)\widehat{p}(y_{1:T}|\theta)^{\lambda_k}, \quad 0 = \lambda_0 < \lambda_1 < \cdots < \lambda_K = 1. \tag{2.15}$$

This approach is particularly robust to multimodal posteriors and outliers, as the tempering schedule allows for gradual exploration of the parameter space.

### 2.4.3 Computational Challenges and Variance Reduction

The asymptotic efficiency of both PMMH and SMC$^2$ is strictly governed by the variance of the log-likelihood estimator, $\mathrm{Var}[\log \widehat{p}_\theta(y_{1:T})]$. Theoretical consistency results suggest targeting a standard deviation of approximately 1 to ensure optimal mixing regimes (Pitt et al., 2012; Sherlock et al., 2015). If the variance exceeds this threshold, the algorithm frequently samples spuriously high likelihood values; these high likelihood estimates enter the denominator of the Metropolis-Hastings acceptance ratio (2.13), permanently freezing the chain and precluding convergence. Achieving this stability requires overcoming two fundamental barriers, which are direct reflections of the filtering challenges identified in Section 2.3.

The first barrier is the informative observation problem inherited from the underlying state-space model. The challenge of highly informative observations manifests in the pseudo-marginal context as a rapid accumulation in the variance of the likelihood estimator. As established previously, simple proposals (such as the BPF) yield negligible ESS in these regimes. In parameter inference, this inefficiency translates directly into a noisy, high-variance likelihood estimate that destabilizes the MCMC chain. The deployment of

the advanced filtering variants discussed in Section 2.3 is indispensable.

The second barrier is the curse of time. Even with an optimal filter, the variance of the log-likelihood estimator accumulates linearly with the time horizon $T$. For long financial time series ($T > 1000$), the intrinsic Monte Carlo noise can still hinder the acceptance of small, beneficial parameter moves. One effective way to mitigate this without a prohibitive increase in the particle count $N$ is to use CRNs. By fixing the seed of the random number generator used by the inner particle filter, we induce a strong positive correlation between the likelihood estimates of the current parameter $\theta^{(i-1)}$ and the proposed $\theta^*$. This correlation ensures that the noise components in the acceptance ratio numerator and denominator largely cancel out, isolating the true likelihood difference and significantly improving mixing efficiency, as rigorously established by Deligiannidis et al. (2018). Note, however, that the application in Chapter 3 departs slightly from this canonical formulation. There, we employ a modified CRN scheme specifically designed to exploit the continuous structure of our smooth likelihood approximation.

### 2.4.4 Applications in Macroeconomics and Finance

Pseudo-marginal methods have fundamentally altered the landscape of empirical macroeconomics and finance. By enabling exact Bayesian inference for nonlinear, non-Gaussian state-space models, these tools allow researchers to estimate structural parameters directly from data without relying on the restrictive linear approximations or moment-matching techniques that characterized earlier literature.

**Bayesian Inference on DSGE Models** Dynamic stochastic general equilibrium (DSGE) models serve as one of the foundational components of modern macroeconomic policy analysis. Historically, Bayesian estimation relied heavily on the Kalman Filter applied to linearised versions of the model (An and Schorfheide, 2007). While effective for small shocks, this approach fails to capture asymmetries, time-varying volatility, or the effects of the zero lower bound on interest rates. The introduction of SMC methods provided a significant advancement. Herbst and Schorfheide (2015) demonstrated that SMC samplers could reliably estimate nonlinear DSGE models intractable for standard MH algorithms. More recently, Fulop et al. (2022b) advanced this frontier by combining controlled SMC by Heng et al., 2020 with an adaptive SMC$^2$ algorithm to achieve numerically stable likelihood

estimates for New Keynesian models with stochastic volatility.

**Bayesian Inference on Long-Run Risk Models** In asset pricing, LRR models (Bansal and Yaron, 2004) posit that asset prices are driven by a small, persistent component in consumption growth and time-varying economic uncertainty. The shift toward full-information Bayesian inference was pioneered by Schorfheide et al. (2018), who developed a Bayesian mixed-frequency state-space model estimated via Particle MCMC using a specialized Rao-Blackwellised particle filter. However, their reliance on log-linearisation can introduce approximation errors regarding non-negative volatility processes. Fulop et al. (2022a) addressed these limitations by implementing a fully nonlinear SMC$^2$ approach. Unlike previous work, their method used autoregressive gamma processes for volatility and projection methods for model solution, demonstrating that rigorous nonlinear estimation is essential for disentangling the dynamics of long-run consumption risks.

## 2.5 SMC for Optimization and Selection

Beyond sequential state estimation, the generic SMC framework adapts to target static distributions, serving as a versatile tool for global optimization (Duan et al., 2023).

The fundamental approach involves reformulating optimization as a sampling problem. An optimization objective, $\max_{x \in \mathsf{X}} h(x)$, is transformed into finding the mode of a non-negative target function, $\gamma_T(x) \propto \exp(h(x))$, which is treated as a density up to an unknown normalizing constant. The SMC estimate for the optimum is the particle $x^{(i)}$ in the final sample that yields the highest functional value $\gamma_T(x^{(i)})$.

### 2.5.1 The Tempered SMC Sampler

To apply the generic backbone, we define the target sequence not by data arrival, but by a tempering schedule. We construct a path bridging a simple initialization sampler $\pi_0$ (where particles are easily drawn) to the complex objective $\pi_T = \gamma_T$.

We introduce an artificial inverse temperature parameter $\lambda_t \in [0, 1]$, such that $0 = \lambda_0 <$

$\lambda_1 < \cdots < \lambda_T = 1$. The sequence of intermediate densities, $\gamma_t(x)$, is defined as:

$$\gamma_t(x) = \gamma_T(x)^{\lambda_t}\, \pi_0(x)^{1-\lambda_t}, \qquad t = 0, \ldots, T. \tag{2.16}$$

At step $t = 0$, the target is the simple $\pi_0$. At step $t = T$, the target is the objective $\gamma_T$.

At each iteration $t$, particles $\{x_{t-1}^{(i)}\}_{i=1}^N$ are reweighted using the incremental weight derived from the ratio of successive densities:

$$w_t(x_{t-1}^{(i)}) = \frac{\gamma_t(x_{t-1}^{(i)})}{\gamma_{t-1}(x_{t-1}^{(i)})} = \left[\frac{\gamma_T(x_{t-1}^{(i)})}{\pi_0(x_{t-1}^{(i)})}\right]^{\lambda_t - \lambda_{t-1}}. \tag{2.17}$$

This process introduces the complexity of the objective function gradually (Neal, 2001). Unlike simulated annealing which operates on a single chain, Tempered SMC operates on a population. The resampling step allows the algorithm to discard particles trapped in poor local optima and reallocate computational effort to promising regions, ensuring robustness in multimodal landscapes.

The choice of the temperature schedule $\{\lambda_t\}$ is crucial for stability. Adaptive tempering schemes choose the next temperature $\lambda_t$ dynamically to achieve a target ESS reduction (e.g., ensuring $\text{ESS}_t \approx 0.5\text{ESS}_{t-1}$), making the schedule responsive to the complexity of the path (Zhou et al., 2016).

### 2.5.2 The Rejuvenation Step: Boosting Particle Diversity

The rejuvenation step is critical in optimization to counteract particle degeneracy and facilitate exploration. It applies a kernel $K_t$ (typically MH or MALA) to move particles within the current target distribution $\pi_{\lambda_t}$. Unlike in filtering, where the goal is simply to avoid degeneracy, in optimization the number of MCMC steps ($R$) is often increased to ensure particles can traverse the energy landscape and escape local optima. A mixture proposal combining independent global jumps and local random walks is often effective for this purpose (Duan et al., 2023).

### 2.5.3 Application to Discrete and Combinatorial Spaces

Tempered SMC is particularly suitable for discrete optimization, such as high-dimensional variable selection (Duan, 2019). Here, the state-space $\mathsf{X}$ is the set of all possible predictor combinations. A particle $x$ represents a specific subset of predictors. For linear regression, the target is defined by the Residual Sum of Squares:

$$\gamma_T(x) \propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - X_x\hat{\beta}_x\|_2^2\right). \tag{2.18}$$

The rejuvenation kernel utilizes discrete transition mechanisms, such as the random substitution of a currently selected predictor for an unselected candidate. Tempering flattens the multimodal landscape at low $\lambda_t$, allowing the sampler to escape suboptimal variable combinations that would trap a standard greedy algorithm.

## 2.6 Conclusion: Bridging Theory and Application

This chapter has established the unified Sequential Monte Carlo framework that constitutes the methodological backbone of this thesis. By formally varying the interpretation of the sequence index $t$ and the sequence of target distributions $\pi_t$, we have delineated two distinct algorithmic paradigms: the Path of Partial Posteriors for dynamic time-series analysis and the geometric path for static inference and optimization. The theoretical framework constructed at the beginning explicitly underpins the empirical contributions presented in the subsequent chapters.

First, mapping the generic framework to the path of partial posteriors yields the standard particle filtering algorithms. By defining the target sequence as the evolving posterior $\pi_t = p_\theta(x_{0:t}|y_{1:t})$, the generic forward kernel $M_t$ specializes into the proposal density $q_t(x_t|x_{t-1}, y_t)$. This derivation provides the rigorous foundation for the unscented particle filter with smooth likelihood utilized in Chapter 3. Furthermore, we established that the computational viability of parameter inference is strictly dependent on the variance of the unbiased likelihood estimator $\widehat{p}_\theta(y_{1:T})$ produced by this filter. Consequently, overcoming the curse of time requires the variance reduction techniques, specifically CRNs and guided proposals, that are central to the Long-Run Risk application.

Conversely, by replacing the physical time index with an artificial inverse temperature schedule $\lambda_t$, we mapped the generic framework to the geometric path. In this context, the sequence of distributions bridges a tractable reference measure to a complex posterior or objective function. This formulation is pivotal for the two distinct applications addressed in the remainder of this work. Specifically, for continuous parameter inference in Chapter 3, we employ density tempering within an SMC$^2$ algorithm to robustly estimate the static structural parameters $\theta$ of the nonlinear LRR model. In a similar vein, for the discrete combinatorial optimization in Chapter 4, we adapt the same geometric principle to probability tempering, which enables the efficient exploration of the high-dimensional discrete search space inherent to the Credit Recovery variable selection problem.

In summary, while the applications in Chapters 3 and 4 appear distinct—ranging from macro-finance latent state estimation to credit risk feature selection—they are mathematically unified instances of the single generic SMC framework established here.

# Chapter 3

# Bayesian Parameter Estimation in Long-Run Risk Models with Mixed-Frequency Data

**Authors.** Andras Fulop, Jeremy Heng, Junye Li, Patrick Tang.

**Abstract.** This paper develops a Bayesian inference framework for estimating long-run risk models with nonlinear state dynamics and mixed-frequency macro-financial data. Standard methods such as Particle Markov chain Monte Carlo (PMCMC) and SMC$^2$ require unbiased likelihood estimates and become computationally infeasible in settings with highly informative observations and latent volatility. We propose an alternative estimation strategy, S-UPF-CRN-SMC, that replaces the pseudo-marginal requirement with a smooth likelihood approximation based on Gaussian filtering and common random numbers (CRNs). This method induces strong correlation across likelihood evaluations, stabilizes posterior inference, and reduces runtime by several orders of magnitude. We apply the framework to a long-run risk model with stochastic volatility and preference shocks, estimated on U.S. data from 1959 to 2022. The use of mixed-frequency data—quarterly consumption and dividend growth, alongside monthly returns and interest rates—enables identification of short-run risk components that are obscured in temporally aggregated samples. Our results show that the inclusion of the COVID-19 period leads to marked reductions in estimated risk aversion, volatility persistence, and dividend growth dynamics, consistent with a transitory shock interpretation. The findings highlight both the robustness

of the proposed method and the empirical value of high-frequency data for identifying temporary shifts in macro-financial risk.

## 3.1 Introduction

This paper develops a Bayesian estimation framework for long-run risk (LRR) models that is both computationally efficient and robust to nonlinearities and mixed-frequency observations. We introduce two key innovations: first, a smooth particle filter that uses a Gaussian approximation to stabilize the likelihood surface; and second, the use of common random numbers (CRNs) to induce correlation across parameter draws, enabling efficient posterior sampling without relying on standard pseudo-marginal methods. By circumventing the instability and high computational cost of likelihood-based estimation in highly nonlinear and mixed-frequency settings, our approach offers a fast and numerically stable alternative for inference in long-run risk models. Empirically, we apply the method to U.S. asset market data—including the COVID-19 period—and demonstrate that our approach delivers sharper posterior distributions and meaningful economic insights while reducing runtime by several orders of magnitude.

The LRR model of asset pricing, developed by Bansal and Yaron (2004), has been instrumental in addressing the equity premium puzzle. The model posits that consumption and dividend growth contain slowly evolving components that, while statistically subtle, have significant implications for asset prices. When embedded within recursive preference structures, such as those proposed by Epstein and Zin (1989) and Weil (1989), the LRR framework accounts for observed asset return premia and macro-finance dynamics through the interaction of long-run growth risk, time-varying volatility, and investor preferences.

Despite its theoretical appeal, the LRR model poses significant challenges for empirical implementation, particularly in estimating parameters that govern persistent growth and volatility dynamics. Several studies have developed simulation-based estimation strategies to address these challenges, including indirect inference, particle filtering, and Bayesian Sequential Monte Carlo (SMC) methods. Our work builds on this foundation but introduces key innovations in how it handles nonlinearity, latent state dynamics, and mixed-frequency data.

Calvet and Czellar (2015) propose a related simulation-based approach for estimating

the long-run risk model. They use indirect inference for parameter estimation, since the full structural likelihood can be highly irregular and difficult to evaluate directly. Their method relies on a simpler auxiliary equilibrium for which the latent state can be recovered and the likelihood is available in closed form. In addition, they use Approximate Bayesian Computation (ABC) filtering to extract information about the hidden states and to produce conditional forecasts. Their paper therefore shows that simulation-based methods can make both parameter estimation and state filtering feasible in long-run risk models. Their approach is nevertheless distinct from the Bayesian particle-filtering methods considered here, since parameter estimation is based on matching auxiliary equilibrium statistics rather than constructing a posterior distribution through pseudo-marginal or SMC-based likelihood estimation.

Schorfheide et al. (2018) introduce a mixed-frequency LRR model with separate volatility processes for consumption growth, the long-run risk component, and dividend growth, using log-normal specifications. While they accommodate mixed-frequency observations by assuming latent dynamics evolve monthly, their reliance on linearisation techniques to approximate asset prices introduces discrepancies between the data-generating process and the pricing implications. Moreover, they assume fixed measurement error variance for risk-free rates, which complicates the interpretation of their asset pricing implications. Their framework, though pioneering, is ultimately limited by its reliance on a linearised model and a Rao-Blackwellised particle filter, both of which constrain its ability to capture nonlinear interactions between state variables and observables.

In contrast, Fulop et al. (2022a) explore LRR models using autoregressive (AR) and autoregressive gamma (ARG) processes for modelling stochastic volatility. Their results demonstrate that ARG-based models improve fit relative to traditional LRR specifications (Bansal and Yaron, 2004; Bansal et al., 2007; Bansal et al., 2016). However, their model omits preference shocks—an important feature in macro-finance applications—and remains limited to balanced-frequency data with quarterly observations. While Fulop et al. (2024) extend the framework to include preference shocks and nonlinear volatility structures using $\text{SMC}^2$ with unscented Kalman filter (UKF) proposals, their setting is also restricted to balanced-frequency designs. None of these approaches provide a computationally viable strategy for posterior inference in highly nonlinear LRR models with mixed-frequency data and informative observations.

The Bayesian branch of the recent literature, including Schorfheide et al. (2018), Fulop

et al. (2022a) and Fulop et al. (2024), estimates LRR models using Particle Markov Chain Monte Carlo (PMCMC) or SMC$^2$ methods. These approaches adopt a two-layer structure: an inner particle filter integrates out latent states conditional on parameters, while an outer sampler—either MCMC or SMC—targets the posterior distribution over parameters. Both rely on approximating the intractable likelihood using Monte Carlo methods and require careful control of variance to ensure stable performance.

While these techniques are theoretically sound, they become computationally infeasible in complex models with nonlinear state dynamics and highly informative observations. In our setting, using a standard SMC$^2$ approach with the Unscented Particle Filter (UPF) would require over 400,000 likelihood evaluations. Even with 14-core parallelization, this leads to runtimes exceeding 200 days. Moreover, the variance of the log-likelihood estimator remains high—particularly in mixed-frequency models—causing poor mixing and unstable posterior inference. These challenges underscore the practical limitations of pseudo-marginal methods when applied to real-world macro-finance datasets.

To overcome these barriers, we introduce a computationally efficient alternative that circumvents the pseudo-marginal framework altogether. Rather than requiring unbiasedness, we fix the random inputs used in the likelihood estimator, by employing CRNs across parameter draws, to induce strong positive correlation in the resulting estimates. In addition, we replace the discrete particle-based likelihood with a smoothed approximation based on a Gaussian density centered on the estimated state mean. This approach reduces Monte Carlo noise, stabilizes Metropolis-Hastings acceptance probabilities, and dramatically improves computational efficiency. Although this method introduces bias, it preserves the curvature of the likelihood surface and enables fast and robust posterior inference in settings that are otherwise intractable with conventional techniques. We further validate the accuracy of this framework in synthetic Monte Carlo experiments, where the method successfully recovers known parameters with low root mean square error, confirming its statistical integrity.

We apply this framework to a long-run risk model with stochastic volatility and preference shocks, estimated using a mixed-frequency dataset in which consumption and dividend growth are observed quarterly, while risk-free rates and equity returns are available monthly. This structure allows us to model latent state dynamics at the monthly level while comparing the empirical implications of balanced- and mixed-frequency data within a unified framework.

Our findings show that mixed-frequency estimation delivers sharper posterior distribu-

tions and more responsive parameter inference, particularly when the COVID-19 period is included. Compared to balanced-frequency models, the mixed-frequency specification yields a substantially lower posterior mean for the risk aversion coefficient, a noticeable drop in volatility persistence, and weaker dividend growth persistence. At the same time, the core long-run primitives—discounting, intertemporal substitution, and the persistence of trend growth—remain essentially unchanged across samples. These patterns are consistent with the interpretation of the pandemic as a transitory risk shock that temporarily altered investor beliefs about volatility and payouts without undermining long-run growth expectations. Consistent with this reading, the short-run adjustment component reacts sharply at recession onsets, the volatility factor jumps during the Global Financial Crisis in 2008 and COVID-19, and the model's predictive bands widen precisely in those episodes.

Our results also reinforce the argument advanced by Schorfheide et al. (2018) that high-frequency data can improve inference about persistent components. However, while their framework relies on a linearised model and is limited to pre-COVID data, we extend this insight by showing that mixed-frequency modelling improves posterior precision even under structural breaks and in fully nonlinear settings. Methodologically, our particle-filtering pipeline with CRNs stabilizes inference and yields tight, interpretable posteriors; in Monte Carlo it produces clear posterior shrinkage relative to balanced-frequency designs, and in the data it delivers predictive bands that widen in real time during crises. The posterior distributions we obtain are not only tighter but also more consistent with observed crisis-period asset price behaviour. Taken together, these features highlight the empirical value of incorporating high-frequency data when estimating long-run risk models and the practical advantage for real-time monitoring: short-run turbulence is detected and quantified without distorting long-run beliefs.

The remainder of the paper is organized as follows. Section 2 presents the long-run risk model, including its recursive preferences and stochastic volatility structure. Section 3 reformulates the model in a state-space representation that accommodates mixed-frequency data and latent monthly dynamics. Section 4 introduces our econometric framework, detailing the smooth particle filtering method, the use of CRNs, and our rationale for departing from pseudo-marginal inference. It also presents an empirical validation of the econometric framework, to step by step showcase the advantages of our proposal relative to standard methods. Section 5 presents the empirical results using U.S. data from 1959 to 2022, comparing parameter estimates across data frequencies and sample compositions—both

including and excluding the COVID-19 period. This section highlights how the mixed-frequency model enhances inference and captures transitory risk dynamics during crises. Finally, Section 6 concludes, summarizing the main contributions and outlining avenues for future research, including nonlinear solution techniques and international extensions of the empirical analysis.

## 3.2 A Consumption-based Long-Run Risk Model

We examine an endowment economy in which a representative agent who has recursive preferences as in Epstein and Zin (1989) and Weil (1989). We follow Albuquerque et al. (2016) and Schorfheide et al. (2018) and assume time preference shocks in the utility function. As shown by Albuquerque et al. (2016), a major role of time preference shocks is to mitigate the strong correlation between stock returns and measurable fundamentals that is typically obtained in consumption-based models without demand shocks. The agent's recursive utility function is given by

$$V_t = \left[ (1-\delta)\lambda_t C_t^{\frac{1-\gamma}{\theta^*}} + \delta \left[ \mathbb{E}(V_{t+1}^{1-\gamma}) \right]^{\frac{1}{\theta^*}} \right]^{\frac{\theta^*}{1-\gamma}}, \tag{3.1}$$

where $C_t$ is consumption at time $t, 0 < \delta < 1$ is the agent's time preference parameter, $\lambda_t$ is the preference shock to the time rate of preference, $\gamma$ is the relative risk aversion (RRA) parameter, $\psi$ is the elasticity of intertemporal substitution (EIS), $\theta^* = \frac{1-\gamma}{1-1/\psi}$ and $\mathbb{E}_t$ stands for conditional expectation with respect to information up to time $t$. This class of preferences allows for a separation between risk aversion and the EIS. The agent prefers early (late) resolution of uncertainty when $\gamma > 1/\psi$ ($\gamma < 1/\psi$). When $\gamma > 1/\psi$, the agent has constant relative risk aversion (CRRA) preferences and is neutral to the timing of resolution of uncertainty. The agent's utility maximization is subject to the following budget constraint,

$$W_{t+1} = (W_t - C_t)R_{t+1}^W,$$

where $W_t$ is the agent's wealth and $R_t^W$ is the return on the portfolio. For any asset $i$ with ex-dividend price $P_{i,t}$ and dividend $D_{i,t}$ the standard Euler equation holds, i.e.,

$$\mathbb{E}_t[M_{t+1}R_{i,t+1}] = 1, \tag{3.2}$$

where $R_{i,t+1} = (P_{i,t+1} + D_{i,t+1})/P_{i,t}$ and $M_t$ is the stochastic discount factor. In particular, for the risk-free asset, we have $R_{f,t} = 1/\mathbb{E}_t[M_{t+1}]$. The stochastic discount factor (SDF) for the recursive utility function defined in (3.1) is given by

$$M_{t+1} = \delta \frac{\lambda_{t+1}}{\lambda_t} \left( \frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \left( \frac{V_{t+1}}{[\mathbb{E}_t(V_{t+1}^{1-\gamma})]^{\frac{1}{1-\gamma}}} \right)^{\frac{1}{\psi}-\gamma}, \tag{3.3}$$

which can be alternatively expressed as

$$M_{t+1} = \delta^{\theta^*} \frac{\lambda_{t+1}}{\lambda_t} \left( \frac{C_{t+1}}{C_t} \right)^{-\frac{\theta^*}{\psi}} (R_{t+1}^W)^{\theta^*-1}. \tag{3.4}$$

Thus, Equation (3.2) implies that for the return on the wealth portfolio,

$$\mathbb{E}_t \left[ \delta^{\theta^*} \frac{\lambda_{t+1}}{\lambda_t} \left( \frac{C_{t+1}}{C_t} \right)^{-\frac{\theta^*}{\psi}} (R_{t+1}^W)^{\theta^*} \right] = 1 \tag{3.5}$$

and for the return on the market portfolio, $R_{m,t}$, we have

$$\mathbb{E}_t \left[ \delta^{\theta^*} \frac{\lambda_{t+1}}{\lambda_t} \left( \frac{C_{t+1}}{C_t} \right)^{-\frac{\theta^*}{\psi}} (R_{t+1}^W)^{\theta^*-1} R_{m,t+1} \right] = 1. \tag{3.6}$$

### 3.2.1 Fundamentals

As in Albuquerque et al. (2016) and Schorfheide et al. (2018), the growth rate of the preference shock, denoted by $x_{\lambda,t} \equiv \ln(\lambda_t/\lambda_{t-1})$, follows the following AR(1) process

$$x_{\lambda,t} = \rho_\lambda x_{\lambda,t-1} + \sigma_\lambda \eta_{\lambda,t+1},$$

in which the shocks to the growth rate of the preference shock are independent of the shocks to consumption growth and dividend growth.

We follow Bansal and Yaron (2004) and Bansal et al. (2016) and assume that the log-consumption growth, $\Delta c_t \equiv \ln\left(\frac{C_t}{C_{t-1}}\right)$ consists of a persistent component, $x_t$, and a

transitory component

$$\Delta c_t = \mu + x_t + \sigma_t \eta_{c,t},$$
$$x_t = \rho x_{t-1} + \phi_x \sigma_t \eta_{x,t},$$

and that dividends are imperfectly correlated with consumption and their log-growth rate, $\Delta d_t \equiv \ln\left(\frac{D_t}{D_{t-1}}\right)$, has the following dynamics,

$$\Delta d_t = \mu_d + \Phi x_{t-1} + \phi_{dc}\sigma_{t-1}\eta_{c,t} + \phi_d \sigma_{t-1}\eta_{d,t},$$

where $\eta_{c,t}, \eta_{d,t}, \eta_{x,t}$ are i.i.d normal $N(0,1)$, and $\sigma_t^2$ is the conditional variance of consumption growth.

The standard long-run risk model Bansal and Yaron (2004), Bansal et al. (2007), and Bansal et al. (2016) assumes that the conditional variance $\sigma_t^2$ follows an autoregressive process of order 1 (AR(1)). However, this modelling choice implies that negative values of $\sigma_t^2$ are possible, which renders the log-linear solution problematic. In addition, Fulop et al. (2022a) show that estimation of long-run risk models with an ARG process of conditional variance, proposed by Gourieroux and Jasiak (2006), performs better than that with the AR(1) process in fitting the market data [1]. To this end, we follow Fulop et al. (2022a) to model the conditional variance as an ARG process with order 1:

$$\sigma_t^2 \sim Gamma(\phi_s + \epsilon_t, c), \quad \epsilon_t \sim Poisson(\frac{\nu\sigma_{t-1}^2}{c})$$

where $Gamma(\cdot)$ and $Poisson(\cdot)$ denote the gamma distribution and the Poisson distribution, respectively, $\nu$ controls the persistence, $c$ determines the scale, and to ensure positivity of conditional variances, the Feller condition, $\phi_s > 1$, needs to be satisfied. The stationary distribution of the ARG process is $Gamma(\phi_s, c/(1-\nu))$ with the long-run mean given by $\bar{\sigma}^2 = \phi_s c/(1-\nu)$.

[1] Zviadadze (2017) also model consumption variance as the ARG(1) process in studying the term structure of currency returns.

## 3.3 Long-Run Risk Model as State-Space Model

State-space models establish a flexible and easily interpretable framework to handle time series (Doucet and Johansen, 2009). For time $t = 1, \ldots, T$, $y_t$ denotes a vector of observations we have at time step $t$ and $y_t$ are conditionally independent given the hidden Markov process $\{s_t\}_{t=0}^{T}$, where $s_t$ is a vector of latent state variables, including both endogenous and exogenous variables. Note that the state $s_t$ and observations $y_t$ may both be multidimensional. Formally, the state-space model is described by a state transition density:

$$s_0 \sim \mu_\theta(s_0), \quad s_t \sim f_\theta(\cdot|s_{t-1}), \quad t = 1, \ldots, T, \tag{3.7}$$

and an observation density:

$$y_t \sim g_\theta(\cdot|s_t, s_{t-1}), \quad t = 1, \ldots, T, \tag{3.8}$$

which depend on the parameters $\theta$. In this work, only the homogeneous case is considered, such that $f_\theta(\cdot)$ and $g_\theta(\cdot)$ are independent of the time index.

### 3.3.1 Re-write Long-Run Risk Model

The long-run risk model can be cast into the framework of nonlinear and non-Gaussian state-space models. Let us first introduce the scenario with data obtainable at the same frequency. $z_t = \{x_{\lambda,t}, x_t, \sigma_t^2\}$ are the standard latent variables of the LRR model as described in Section 3.2, where $x_{\lambda,t}$ is the growth rate of the preference shock, and $x_t$ and $\sigma_t^2$ are the long-run component of consumption growth and its conditional variance. We have:

$$\begin{aligned} x_t &= \rho x_{t-1} + \phi_x \sigma_t \eta_{x,t}, \\ x_{\lambda,t} &= \rho_\lambda x_{\lambda,t-1} + \sigma_\lambda \eta_{\lambda,t}, \\ \sigma_t^2 &\sim \text{Gamma}(\phi_s + \epsilon_t, c), \quad \epsilon_t \sim \text{Poisson}\left(\frac{\nu \sigma_{t-1}^2}{c}\right), \end{aligned} \tag{3.9}$$

where $\eta_{x,t}, \eta_{\lambda,t} \sim \mathcal{N}(0,1)$ are independent noise terms. Both the growth rate and the preference shock $x_\lambda$ are modelled as AR(1) processes, whereas the evolution of the con-

sumption volatility, $f_\theta(\sigma_t^2|\sigma_{t-1}^2)$, follows an ARG process and its marginal transition kernel is a non-central Gamma distribution. The exact derivation can be found in the Appendix. By initializing the latent process at $z_0 = (x_0, x_{\lambda,0}, \sigma_0^2) = (0, 0, \bar{\sigma}^2)$, the Markov transition kernel of the latent process has the form:

$$f_\theta(z_t|z_{t-1}) = f_\theta(x_t, x_{\lambda,t}|x_{t-1}, x_{\lambda,t-1}) f_\theta(\sigma_t^2|\sigma_{t-1}^2), \tag{3.10}$$

where

$$f_\theta(x_t, x_{\lambda,t}|x_{t-1}, x_{\lambda,t-1}) = \mathcal{N}\left(x_t; \rho x_{t-1}, \phi_x^2 \sigma_t^2\right) \times \mathcal{N}\left(x_{\lambda,t}; \rho_\lambda x_{\lambda,t-1}, \sigma_\lambda^2\right).$$

There are four observables including the consumption growth rates ($\Delta c_t$), the dividend growth rates ($\Delta d_t$), the market returns ($r_{m,t}$), and the risk-free returns ($r_{f,t}$). Assuming that market and risk-free returns are collected with measurement errors, the dynamics of the measurements is given by:

$$\begin{aligned}
\Delta d_t &= \mu_d + \Phi x_{t-1} + \phi_{dc}\sigma_{t-1}\eta_{c,t} + \phi_d \sigma_{t-1}\eta_{d,t}, \\
r_{m,t} &= \xi_\theta(z_t, z_{t-1}, \Delta d_t) + \phi_m \eta_{m,t}, \\
r_{f,t} &= \zeta_\theta(z_t) + \phi_f \eta_{f,t}, \\
\Delta c_t &= \mu + x_t + \sigma_t \eta_{c,t},
\end{aligned} \tag{3.11}$$

where $\eta_{c,t}, \eta_{d,t}, \eta_{m,t}, \eta_{f,t} \sim \mathcal{N}(0,1)$. The conditional mean of market return $\xi_\theta(z_t, z_{t-1}, \Delta d_t)$ is given by:

$$\xi_\theta(z_t, z_{t-1}, \Delta d_t) = \log\{\exp[PD_\theta(z_t)] + 1\} - PD_\theta(z_{t-1}) + \Delta d_t,$$

where $PD(z_t, \sigma_t^2)$ denotes the log-price-dividend ratio. The conditional mean of $\zeta_\theta(z_t)$ has the form:

$$\zeta_\theta(z_t) = -PF_\theta(z_t),$$

where $PF(z_t, \sigma_t^2)$ represents the log-wealth-consumption ratio.

The usual method to solve the long-run risk models is the log-linear approximation method of Campbell and Shiller (1988); see Bansal and Yaron (2004), Bansal et al. (2016), Beeler and Campbell (2009), Schorfheide et al. (2018), among others. In the current paper, we adopt log-linear solutions to the long-run risk models with preference shocks and the ARG volatility process, as discussed in Fulop et al. (2024).

### 3.3.2 Model with Mixed-Frequency Data

The standard model assumes all four observations are available at the same frequency, such that at each time point all necessary data are monthly observable. We want to focus on the special case, where the consumption and the dividend are quarterly observable and the other two observations are monthly available. Schorfheide et al. (2018) define quarterly consumption as the sum of monthly consumption over the span of five months and the quarterly consumption growth rates are given by, w.r.t. the time index $t$:

$$\begin{aligned}\Delta c_t^q &= \sum_{k=1}^{5} \left( \frac{3 - |k-3|}{3} \right) \Delta c_{t-k+1} \\ &= \frac{1}{3}\Delta c_t + \frac{2}{3}\Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3}\Delta c_{t-3} + \frac{1}{3}\Delta c_{t-4},\end{aligned} \tag{3.12}$$

is observable if and only if $t \equiv 0 \pmod 3$. In simple terms, at the end of a quarter, it is not appropriate to simply sum up these three monthly consumption growth of that quarter, but to take the weighted consumption of the previous five month into account. Especially, it is logical to put more weight on the end month of previous quarter, at which we obtain the last observation. If we use (3.9) and (3.11) directly, it will violate the basic assumption of state-space model, such that the latent variables at time $t$ are only dependent on the latent variables at time $t-1$. Hence, it is necessary to rewrite the state space by introducing more latent variables. Furthermore, it is crucial to notice that, in the standard model (3.11), $\Delta c_t$ and $\Delta d_t$ are intertwined, since they are also sharing the same noise term $\eta_{c,t}$. One possible way to write the state-space is to save the necessary information in the form of four running sums. The reason is that $x_t, \sigma_t^2$ will play a role in both the current quarter and the next quarter. Although this approach only needs to add a few variables to the latent space, deriving the associated noise term in aggregated quarterly consumption is complex.

In this paper, we use the same definition for quarterly observations and we extend the dividend observation as:

$$\Delta d_t^q = \frac{1}{3}\Delta d_t + \frac{2}{3}\Delta d_{t-1} + \Delta d_{t-2} + \frac{2}{3}\Delta d_{t-3} + \frac{1}{3}\Delta d_{t-4}. \tag{3.13}$$

Assume we are currently at the end of quarter at time $t$ and we update the state-space every quarter, although the monthly risk free returns and market returns are available at a monthly rate. Therefore, we have the following set of new observations at the end of next

quarter at $t+3$:

$$y_{t+3} = \left\{ \Delta c^q_{t+3}, \Delta d^q_{t+3}, r_{m,t+3}, r_{f,t+3}, r_{m,t+2}, r_{f,t+2}, r_{m,t+1}, r_{f,t+1} \right\}.$$

In this work, we add the necessary variables into the latent space directly. The number of additional variables is larger than the running sum approach, but a clearer model structure can be established.

### 3.3.3 Extended state space

Since in the weighted sum described above, $\Delta c_t, \Delta c_{t-1}, \Delta d_t, \Delta d_{t-1}$ are unobserved terms that already play a role in the previous quarter, they should be updated and be carried as known terms into the next quarter. Let $z^q_t = \{\Delta c_t, \Delta c_{t-1}, \Delta d_t, \Delta d_{t-1}\}$ represent the extended latent variables. Hence, in the full extended space, we have the following set of seven state variables to be updated after each iteration:

$$s_t = \{z_t, z^q_t\} = \{x_{\lambda,t}, x_t, \sigma^2_t, \Delta c_t, \Delta c_{t-1}, \Delta d_t, \Delta d_{t-1}\}.$$

Although the latent space is seven-dimensional, in order to derive the observation density for $y_{t+3}$, the following set of latent variables are necessary:

$$\begin{aligned} &\{z^q_t, z_t, z_{t+1}, z_{t+2}, z_{t+3}\} \\ =&\{\Delta c_t, \Delta c_{t-1}, \Delta d_t, \Delta d_{t-1}, x_{\lambda,t}, x_t, \sigma^2_t, x_{\lambda,t+1}, x_{t+1}, \sigma^2_{t+1}, x_{\lambda,t+2}, x_{t+2}, \sigma^2_{t+2}, x_{\lambda,t+3}, x_{t+3}, \sigma^2_{t+3}\}. \end{aligned}$$

The evolution of $\{z_{t+1}, z_{t+2}, z_{t+3}\}$ is described by:

$$\bar{f}(z_{t+1}, z_{t+2}, z_{t+3}|z_t) = f(z_{t+3}|z_{t+2}) f(z_{t+2}|z_{t+1}) f(z_{t+1}|z_t).$$

How to update the extended latent states, $z^q_{t+3} = \{\Delta c_{t+2}, \Delta c_{t+3}, \Delta d_{t+2}, \Delta d_{t+3}\}$, conditioned on the information at the end of quarter

$$\hat{f}(z^q_{t+3}|y_{t+3}, z^q_t, z_t, z_{t+1}, z_{t+2}, z_{t+3})$$

is described in the next subsection. Then, the full transition equation is defined by:

$$\mathbf{F}(s_{t+3}|s_t) = \hat{f}(z^q_{t+3}|y_{t+3}, z^q_t, z_t, z_{t+1}, z_{t+2}, z_{t+3})\bar{f}(z_{t+1}, z_{t+2}, z_{t+3}|z_t). \tag{3.14}$$

**Joint Density and Measurement**

As information of latent dividends already contained in the above monthly observations, we need to update the formulations for the quarterly consumption and the quarterly dividends. Based on (3.12) and (3.13), with the carried-forward latent observation $\Delta d_{t,t-1}$ and $\Delta c_{t,t-1}$ from last quarter, we have:

$$\begin{aligned}\Delta d^q_{t+3} = {} & 2\mu_d + \Phi x^q_{t+3} + \frac{2}{3}\Delta d_t + \frac{1}{3}\Delta d_{t-1} \\ & + \frac{1}{3}(\phi_{dc}\sigma_{t+2}\eta_{c,t+3} + \phi_d\sigma_{t+2}\eta_{d,t+3}) + \frac{2}{3}(\phi_{dc}\sigma_{t+1}\eta_{c,t+2} + \phi_d\sigma_{t+1}\eta_{d,t+2}) \\ & + (\phi_{dc}\sigma_t\eta_{c,t+1} + \phi_d\sigma_t\eta_{d,t+1})\end{aligned}$$

and

$$\begin{aligned}\Delta c^q_{t+3} = {} & 2\mu + x^q_{t+3} + \frac{2}{3}\Delta c_t + \frac{1}{3}\Delta c_{t-1} \\ & + \frac{1}{3}\sigma_{t+2}\eta_{c,t+3} + \frac{2}{3}\sigma_{t+1}\eta_{c,t+2} + \sigma_t\eta_{c,t+1},\end{aligned}$$

where

$$x^q_{t+3} = \frac{1}{3}x_{t+2} + \frac{2}{3}x_{t+1} + x_t.$$

We would like to describe the set containing the extended latent observations, the quarterly observations, the three monthly market return and risk-free rate observations as the following:

$$
\begin{bmatrix}
\Delta c_{t+2} \\ \Delta c_{t+3} \\ \Delta d_{t+2} \\ \Delta d_{t+3} \\ \Delta c^q_{t+3} \\ \Delta d^q_{t+3} \\ r_{m,t+1} \\ r_{m,t+2} \\ r_{m,t+3} \\ r_{f,t+1} \\ r_{f,t+2} \\ r_{f,t+3}
\end{bmatrix}
=
\begin{bmatrix}
\mu + x_{t+1} \\ \mu + x_{t+2} \\ \mu_d + \Phi x_{t+1} \\ \mu_d + \Phi x_{t+2} \\ 2\mu + x^q_{t+3} + \frac{2}{3}\Delta c_t + \frac{1}{3}\Delta c_{t-1} \\ 2\mu_d + \Phi x^q_{t+3} + \frac{2}{3}\Delta d_t + \frac{1}{3}\Delta d_{t-1} \\ \xi_\theta(z_{t+1}, z_t) + \mu_d + \Phi x_t \\ \xi_\theta(z_{t+2}, z_{t+1}) + \mu_d + \Phi x_{t+1} \\ \xi_\theta(z_{t+3}, z_{t+2}) + \mu_d + \Phi x_{t+2} \\ \zeta_\theta(z_t) \\ \zeta_\theta(z_{t+1}) \\ \zeta_\theta(z_{t+2})
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
0 & \sigma_{t+1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_{t+2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \phi_{dc}\sigma_{t+1} & 0 & 0 & \phi_d\sigma_{t+1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \phi_{dc}\sigma_{t+2} & 0 & 0 & \phi_d\sigma_{t+2} & 0 & 0 & 0 & 0 & 0 & 0 \\
\sigma_t & \frac{2}{3}\sigma_{t+1} & \frac{1}{3}\sigma_{t+2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi_{dc}\sigma_t & \frac{2}{3}\phi_{dc}\sigma_{t+1} & \frac{1}{3}\phi_{dc}\sigma_{t+2} & \phi_d\sigma_t & \frac{2}{3}\phi_d\sigma_{t+1} & \frac{1}{3}\phi_d\sigma_{t+2} & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi_{dc}\sigma_t & 0 & 0 & \phi_d\sigma_t & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 & 0 \\
0 & \phi_{dc}\sigma_{t+1} & 0 & 0 & \phi_d\sigma_{t+1} & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 \\
0 & 0 & \phi_{dc}\sigma_{t+2} & 0 & 0 & \phi_d\sigma_{t+2} & 0 & 0 & \phi_m & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_f & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_f & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_f
\end{bmatrix}
\begin{bmatrix}
\eta_{c,t+1} \\ \eta_{c,t+2} \\ \eta_{c,t+3} \\ \eta_{d,t+1} \\ \eta_{d,t+2} \\ \eta_{d,t+3} \\ \eta_{m,t+1} \\ \eta_{m,t+2} \\ \eta_{m,t+3} \\ \eta_{f,t+1} \\ \eta_{f,t+2} \\ \eta_{f,t+3}
\end{bmatrix}, \tag{3.15}
$$

where the noise terms are i.i.d Gaussian noises and we have:

$$
\begin{aligned}
y^{(joint)}_{t+3} &= \left\{ \Delta c_{t+2}, \Delta c_{t+3}, \Delta d_{t+2}, \Delta d_{t+3}, \Delta c^q_{t+3}, \Delta d^q_{t+3}, r_{m,t+1}, r_{m,t+2}, r_{m,t+3}, r_{f,t+1}, r_{f,t+2}, r_{f,t+3} \right\} \\
&= \mu^{(joint)}_{t+3} + L^{(joint)}_{t+3} \eta^{(joint)}_{t+3}
\end{aligned}
$$

where $L^{(joint)}_{t+3}$ is the corresponding loading matrix. In order to calculate the measurement density of the quarterly observations and the inter-correlated market return, we take a part of the system of equation (B.1):

$$
\begin{aligned}
y_{t+3} &= \left\{ \Delta c^q_{t+3}, \Delta d^q_{t+3}, r_{m,t+1}, r_{m,t+2}, r_{m,t+3}, r_{f,t+1}, r_{f,t+2}, r_{f,t+3} \right\} \\
&= \mu^{(observed)}_{t+3} + L^{(observed)}_{t+3} \eta^{(observed)}_{t+3}
\end{aligned}
$$

where

$$\mu_{t+3}^{(observed)} = \mu_{t+3,5:12}^{(joint)}, \quad L_{t+3}^{(observed)} = \{L_{t+3,i,j}^{(joint)} : 5 \leq i, 1 \leq j\} \text{ and } \quad \eta_{t+3}^{(observed)} = \eta_{t+3,5:12}^{(joint)}.$$

We obtain the full observation density:

$$\mathbf{G}(y_{t+3}|z_t^q, z_t, z_{t+1}, z_{t+2}, z_{t+3}) = \mathcal{N}\Big(\mu_{t+3}^{(observed)}, \Sigma_{t+3}^{(observed)}\Big),$$

where the covariance is $\Sigma_{t+3}^{(observed)} = L_{t+3}^{(observed)}(L_{t+3}^{(observed)})^T$.

Next, we want to sample the unobserved monthly variables $z_{t+3}^q$, based on the quarterly observations:

$$z_{t+3}^q = \{\Delta c_{t+3}, \Delta c_{t+2}, \Delta d_{t+3}, \Delta d_{t+2}\}.$$

We split the observation and their corresponding means into two parts, where the observation part $Y_2$ is known:

$$Y_1 = \begin{bmatrix} \Delta c_{t+2} \\ \Delta c_{t+3} \\ \Delta d_{t+2} \\ \Delta d_{t+3} \end{bmatrix}, Y_2 = \begin{bmatrix} \Delta c_{t+3}^q \\ \Delta d_{t+3}^q \\ r_{m,t+1} \\ r_{m,t+2} \\ r_{m,t+3} \end{bmatrix} \text{ and } \mu_1 = \mu_{t+3,1:4}^{(joint)}, \mu_2 = \mu_{t+3,5:9}^{(joint)},$$

and their corresponding covariance is split into four parts:

$$\begin{aligned}
\Sigma_{11} &= \{\Sigma_{t+3,i,j}^{(joint)} : 1 \leq i \leq 4, 1 \leq j \leq 4\}, \\
\Sigma_{12} &= \{\Sigma_{t+3,i,j}^{(joint)} : 1 \leq i \leq 4, 5 \leq j \leq 9\}, \\
\Sigma_{21} &= \{\Sigma_{t+3,i,j}^{(joint)} : 5 \leq i \leq 9, 1 \leq j \leq 4\}, \\
\Sigma_{22} &= \{\Sigma_{t+3,i,j}^{(joint)} : 5 \leq i \leq 9, 5 \leq j \leq 9\}.
\end{aligned}$$

If $Y_2$ is observed, the conditional distribution of $Y_1|Y_2$ is also normally distributed. The conditional mean and conditional covariance matrix are given by:

$$\begin{aligned}
\mu^{(conditional)} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2) \\
\Sigma^{(conditional)} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\end{aligned}$$

Then we have:

$$\hat{f}(z_{t+3}^q | y_{t+3}, z_t^q, z_t, z_{t+1}, z_{t+2}, z_{t+3}) = \mathcal{N}(\mu^{(conditional)}, \Sigma^{(conditional)}).$$

## 3.4 Econometric Framework

We consider a class of nonlinear state-space models, with the aim of computing the joint posterior distribution of model parameters and latent states, $p(\theta, s_{1:T}|y_{1:T})$. This posterior can be factorized as:

$$p(\theta, s_{1:T}|y_{1:T}) = p(s_{1:T}|\theta, y_{1:T}) \cdot p(\theta|y_{1:T}),$$

where $p(s_{1:T}|\theta, y_{1:T})$ addresses the state smoothing problem, and $p(\theta|y_{1:T})$ refers to the parameter inference problem. According to Bayes' rule:

$$p(\theta|y_{1:T}) \propto p(y_{1:T}|\theta) \cdot p(\theta),$$

where $p(y_{1:T}|\theta)$ is the likelihood and $p(\theta)$ the prior.

### 3.4.1 Pseudo-Marginal Methods and Their Limitations

When the likelihood is analytically intractable, particle filters can be employed to produce an unbiased estimate $\hat{p}(y_{1:T}|\theta)$. This forms the foundation of pseudo-marginal MCMC methods (Andrieu and Roberts, 2009; Andrieu et al., 2010), which sample from an extended posterior:

$$\tilde{p}(\theta, u_{1:T}|y_{1:T}),$$

where $u_{1:T}$ denotes the auxiliary random variables generated by the particle filter (e.g., particles and resampling indices). The key property here is that:

$$p(\theta|y_{1:T}) = \int \tilde{p}(\theta, u_{1:T}|y_{1:T})\, du_{1:T},$$

ensuring that marginal inference over $\theta$ remains exact under the assumption of unbiasedness.

However, these methods are highly sensitive to the variance of the estimated likelihood.

As shown by Doucet et al. (2015), for pseudo-marginal inference to mix efficiently, the standard deviation of the log-likelihood estimate should be close to 1. Otherwise, the Markov chain may become sticky or inefficient.

The core source of instability arises from the structure of the Metropolis-Hastings (MH) acceptance probability:

$$\alpha = \min\left(1, \frac{\hat{p}(y_{1:T}|\theta')p(\theta')}{\hat{p}(y_{1:T}|\theta)p(\theta)}\right),$$

where $\theta' = \theta + \Delta$. The decision to accept a move relies on the difference in log-likelihoods:

$$\log \hat{p}(y_{1:T}|\theta') - \log \hat{p}(y_{1:T}|\theta).$$

When $\hat{p}(\cdot)$ is computed using independent random numbers at each $\theta$, this log-difference is dominated by Monte Carlo noise—even when $\theta$ and $\theta'$ are close. Consequently, the log-likelihoods can be uncorrelated, and the resulting volatility causes poor mixing and algorithmic breakdowns.

### 3.4.2 Using Common Random Numbers and Smoothing the Likelihood

To mitigate the volatility in likelihood estimates inherent in pseudo-marginal methods, we depart from the standard framework and fix the randomness used in the likelihood computation. Specifically, we introduce CRNs $v$ across different values of $\theta$, so the estimated likelihood becomes:

$$\hat{p}(y_{1:T}|\theta, v),$$

and its log-likelihood differences are expressed as:

$$\log \hat{p}(y_{1:T}|\theta', v) - \log \hat{p}(y_{1:T}|\theta, v).$$

Assuming the estimator is reasonably smooth in $\theta$, the use of shared $v$ ensures that the Monte Carlo errors are positively correlated, substantially reducing the variance of the log-likelihood difference. This correlation stabilizes the MH acceptance probabilities and improves chain mixing. Though this practice violates the strict pseudo-marginal framework—since we are no longer integrating over $v$—it provides a practical means to

control volatility in the target distribution surface.

Yet, CRNs alone are insufficient. The particle filter's resampling step introduces discrete discontinuities in the likelihood estimate, breaking smoothness even under fixed randomness. Small changes in $\theta$ may lead to entirely different particle sets being selected, causing sudden jumps in the estimated likelihood. Therefore, smoothness is further reinforced using a Gaussian approximation of the filtering density, as proposed by Fulop and Li (2019).

In dynamic asset pricing models, where observations are highly informative about the latent states, the filtered density of the hidden states tends to concentrate sharply. In such settings, the particle approximation $\hat{p}(s_{1:T}|\theta, y_{1:T})$ can be replaced with a Gaussian density centered at the filtered state mean. This yields a smoother and more stable likelihood surface, enabling robust posterior inference.

To summarize, our method does not fall under the class of pseudo-marginal algorithms, which rely on unbiased likelihood estimates by integrating over the randomness. Instead, we fix the random inputs via CRNs and replace the discontinuous filter-based likelihood with a smooth approximation that remains a function of the model parameters. The resulting likelihood estimate is biased but smooth, and crucially, it behaves well across the parameter space.

Our methodological departure is motivated by the fact that, even with efficient proposals such as the UKF, we find that achieving low-variance pseudo-marginal likelihoods at reasonable computational cost is impractical. By contrast, fixing the randomness and smoothing the likelihood—though no longer providing exact posterior inference—results in a much more tractable and numerically stable approach, suitable for high-dimensional state-space models. We note that this differs also from correlated pseudo-marginal methods (e.g., Deligiannidis et al. (2018)), which induce correlation between likelihood estimates across MCMC iterations, but still preserve the unbiasedness property. In practice, however, the discontinuity introduced by the resampling step in a particle filter drastically reduces any correlation between successive likelihood estimates, even if the underlying random numbers are shared. Moreover, when the state space is high-dimensional, a Hilbert-curve ordering used for multidimensional sorting fails to preserve locality effectively, so correlated resampling no longer maintains close alignment between particle clouds. In our setting, we entirely fix the randomness and construct a deterministic approximation of the likelihood, explicitly moving outside the pseudo-marginal framework.

### 3.4.3 Refinements and Practical Implementation

To operationalize the ideas described above, and in particular the shift from pseudo-marginal methods to fixed-randomness and smooth-likelihood inference, we employ a modified particle filtering approach based on UKF-driven proposals. This section outlines how we adapt the particle filter architecture to our model's structure and computational demands.

We first demonstrate that a smooth filter with UKF proposals for each particle is both more stable and computationally efficient than a standard bootstrap particle filter (BPF). Section 3.3 details the redesigned state-space model.

Fulop et al. (2024) explore the case of balanced-frequency data, where the need to update inter-quarter latent states is absent. Their UKF-based proposal includes all latent components—namely, the persistent state $x$, the growth-rate adjustment state $x_\lambda$, and stochastic volatility $\sigma$—within the unscented transform. This ensures the entire latent state vector is informed by the current observation at each time point. However, a direct extension of the UKF-based proposal to our setting—featuring multi-frequency latent updates—proves problematic.

In particular, we find that applying the unscented transform across three time steps (to update inter-quarter and monthly latent components) produces unstable behaviour in Monte Carlo diagnostics such as Effective Sample Size (ESS) and variance of likelihood estimates, even as particle count increases. This instability originates from the strong nonlinearity of the transition equations, especially in the stochastic volatility component. The UKF approximation is effective under near-linear transitions, but here it leads to over-concentrated proposals, impairing the ability of the filter to explore latent trajectories. By contrast, in the balanced-frequency case the UKF approximation performs well: aggregation reduces idiosyncratic noise and avoids overfitting intra-period fluctuations, yielding stable proposals and diagnostics.

To address this, we develop a hybrid filtering scheme. Specifically, the stochastic volatility component $\sigma$ is updated using a bootstrap proposal, leveraging its transition dynamics directly. Conditional on the sampled volatility path, the UKF-based proposal is applied to $x$ and $x_\lambda$ using the unscented transform. This selective use of the UKF—only where state transitions are moderately nonlinear—preserves stability and avoids over-concentration, yielding robust likelihood estimates.

This modified filter serves as the baseline filtering method used for our likelihood evaluations. Full implementation details are given in the Appendix B.4. To reflect key refinements and innovations, we adopt the following naming conventions:

- S-UPF: A Smoothed Unscented Particle Filter, combining the UKF proposal with a Gaussian approximation to improve computational efficiency and reduce Monte Carlo noise.
- S-UPF-CRN: Enhances S-UPF by using CRNs to fix the randomness in likelihood evaluations. This yields positively correlated log-likelihood estimates and significantly improves MCMC mixing.
- S-UPF-CRN-SMC: Embeds S-UPF-CRN into a Sequential Monte Carlo sampler with adaptive tempering, balancing exploration and exploitation while maintaining robustness against particle impoverishment (see pseudocode in Appendix).

This implementation architecture not only stabilizes inference in the presence of strong nonlinearities but also facilitates a coherent departure from pseudo-marginal methodology by emphasizing smoothness, stability, and tractability in the likelihood evaluation pipeline.

### 3.4.4 Empirical Validation of the Inference Methodology

This subsection provides an empirical rationale for departing from standard pseudo-marginal methods, complementing the conceptual framework in Section 3.4. Our goal is to operationalize the inference by demonstrating that standard unbiased filters are computationally intractable in this setting, and that our proposed alternative, S-UPF-CRN, offers a robust solution. We validate the methodology in three steps: first, we document the computational cost and variance instability of standard filters; second, we assess the accuracy of the Gaussian approximation used to smooth the likelihood; and third, we verify that the combination of smoothing and CRNs preserves the geometry of the likelihood surface, enabling stable MCMC inference.

The validation uses the pre-COVID dataset to isolate algorithmic performance under stable conditions. We run a baseline SMC$^2$ with the S-UPF-CRN to generate posterior draws, then recompute likelihood metrics across a range of parameter vectors selected

from the $\{10, 30, 50, 70, 90, 95, 99\}\%$ quantiles of the joint log-likelihood. This ensures the validation spans the relevant high-probability region of the parameter space..

### Computation Time and Variance Instability

Table 3.1 reports the average computational time required to run the BPF, UPF, and our S-UPF, across increasing numbers of particles. The results are averaged over 50 independent runs, using parallelization to account for practical execution delays and computational overhead. This provides a realistic measure of runtime cost under conditions representative of actual implementation.

**Table 3.1:** Average Computational Runtime (in Seconds)

| **Particles** | **BPF** | **S-UPF** | **UPF** |
|---|---|---|---|
| $2^7$ | 8.93 | 5.70 | 3.54 |
| $2^8$ | 3.48 | 7.99 | 15.30 |
| $2^9$ | 6.16 | 11.67 | 14.66 |
| $2^{10}$ | 12.26 | 24.13 | 25.81 |
| $2^{11}$ | 24.70 | 59.81 | 51.85 |
| $2^{12}$ | 67.90 | 138.99 | 120.25 |
| $2^{13}$ | 139.29 | 294.25 | 263.24 |
| $2^{14}$ | 295.22 | 564.49 | 549.77 |

As expected, BPF is the fastest, with minimal computational overhead. Both UPF and S-UPF require more computation due to their use of local linearisation and unscented transforms. Their runtime is very similar, and approximately double that of BPF.

However, we initially considered the BPF as a baseline but found that it suffered from severe weight collapse, consistent with the literature on high-dimensional and informative state-space models. In the pre-COVID sample, the BPF exhibited particle degeneracy even at high particle counts, yielding log-likelihood variance estimates orders of magnitude higher than the threshold required for stable pseudo-marginal inference. Consequently, we report the BPF runtime here for completeness but exclude it from the subsequent visual analysis to focus on the comparison between the viable alternatives: UPF and S-UPF.

Figure 3.1 plots the standard deviation of the estimates, addressing the question of "how many particles are needed" to achieve stability. We exclude the BPF because its standard deviation exceeds $10^5$, rendering it invisible on a reasonable scale. Focusing on the S-UPF

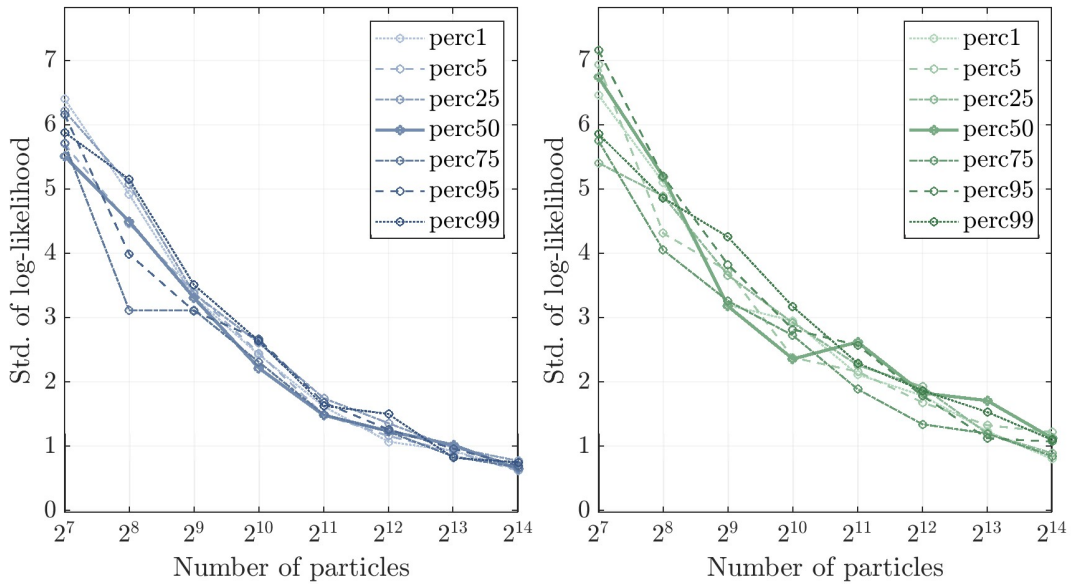



**Figure 3.1:** Standard deviation of log-likelihood estimates for S-UPF (left) and UPF (right) across varying particle counts ($M = 2^7$ to $M = 2^{14}$). Results are averaged over 50 runs using 7 representative parameter sets.

(left) and UPF (right) panels, both filters require a large number of particles to bring the standard deviation down to the target of 1, as recommended by Doucet et al. (2015). Both filters achieve this target at $M = 2^{14}$ particles[2].

The cost of running UPF at scale becomes prohibitive: a single likelihood evaluation takes approximately 10 minutes, or 600 seconds. When embedded in a full SMC$^2$ inference framework—as used in Fulop et al. (2022a) and Fulop et al. (2024)—with 20 tempering steps, 10 MCMC move steps per tempering level, and $2^{11} = 2048$ parameter particles, the total number of likelihood evaluations is

$$2048 \times 10 \times 20 = 409{,}600.$$

This configuration is empirically motivated by early experiments using the BPF on the same dataset, where these settings were necessary to achieve sufficient mixture and adequate exploration of the parameter space. The total computation time, without parallelization, is

$$409{,}600 \times 600 = 245{,}760{,}000 \text{ seconds},$$

[2]By contrast, when the COVID-19 period is included, the standard deviation is close to 10 at $M = 2^{14}$ (see Appendix B.5).

which corresponds to

$$\frac{245{,}760{,}000}{60 \times 60 \times 24} \approx 2{,}844.44 \text{ days.}$$

Our experiments were conducted on a 16-core vCPU machine with 64 GB of RAM—a configuration that reflects a standard mid-tier cloud instance (and exceeds the 4–8 core setups common on personal workstations). To parallelize the 409,600 log-likelihood evaluations efficiently, we launch a MATLAB Parallel Computing Toolbox `parpool` with 14 workers and distribute the outer loop over likelihood evaluations via a `parfor` construct. In a `parfor` loop each iteration is scheduled on one worker, so 14 likelihood calculations run simultaneously, while the remaining two cores are reserved for other SMC$^2$ tasks and system overhead—avoiding CPU oversubscription and ensuring tempering and adaptation steps proceed smoothly. To ensure stability over the multi-day runtime, we implement a periodic restart mechanism for the parallel pool.[3] Even with this 14-core parallelization, the estimated runtime remains over 200 days:

$$\frac{2{,}844.44}{14} \approx 203.17 \text{ days.}$$

In practice, the actual runtime would likely exceed this theoretical lower bound due to the additional computational overhead incurred by the stability-preserving restart mechanisms described above. These constraints render standard pseudo-marginal inference infeasible for our application.

### Gaussian Smoothing and Approximation Accuracy

To address the limitations of standard filtering, we propose the use of a Gaussian approximation to the filtering distribution, forming the basis of S-UPF. This smoothing is essential for convergence but introduces a small bias relative to the standard particle filter. We validate the robustness of this approximation from two complementary perspectives: a global assessment of the statistical bias via the mean log-likelihood, and a local inspection of the distributional fit at specific time points.

[3] Specifically, we shut down and reinitialize the `parpool` after every ten calls to the `parfor` function (roughly every 140 likelihood evaluations). This counteracts a technical issue where long-running `parfor` jobs exhibit unexplained, monotonic increases in memory usage ("memory leaks") on worker nodes. In our tests without this reset, we observed that as RAM consumption approached system limits, the operating system would force the termination of individual worker processes to free up resources. This sequential loss of workers eventually caused the entire parallel construct to collapse, halting the estimation.

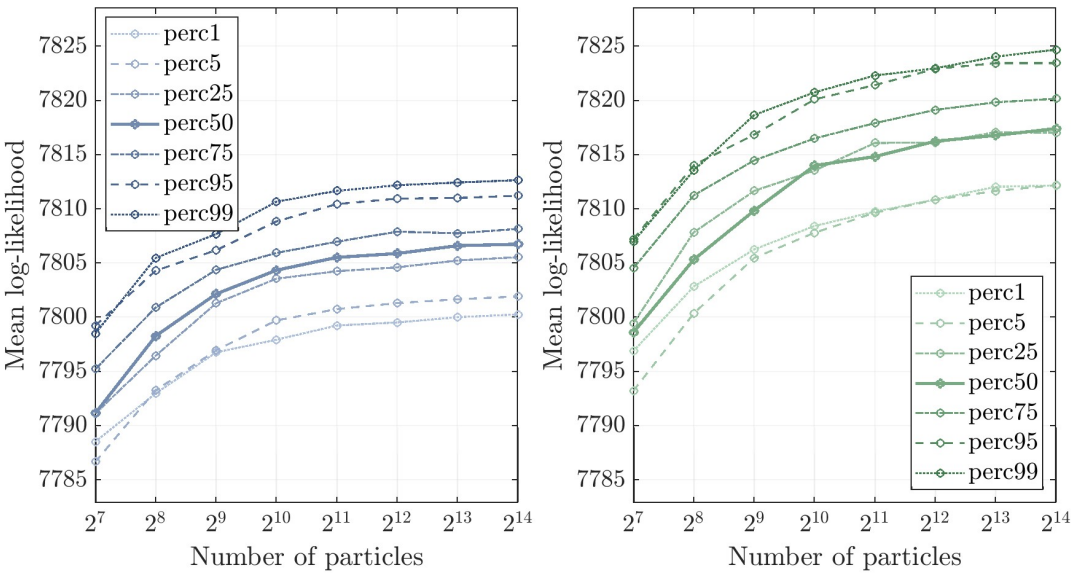



**Figure 3.2:** Mean of log-likelihood estimates for S-UPF (left) and UPF (right) across varying particle counts ($M = 2^8$ to $M = 2^{14}$). Results are averaged over 50 runs using 7 representative parameter sets. The S-UPF shows only a minor downward bias compared to UPF.

Figure 3.2 plots the mean of the log-likelihood estimates to illustrate the global bias quantitatively. As expected, the S-UPF (left), which employs a Gaussian approximation, yields slightly lower likelihood values than the standard UPF (right). However, the difference is small and consistent, confirming that the smoothing does not distort the likelihood surface significantly.

To further validate the approximation at the local level, Figure 3.3 compares the empirical distribution of the filtering distributions during resampling to a normal approximation across seven latent dimensions. The close match across all dimensions supports the validity of the Gaussian assumption and justifies its use to smooth the likelihood surface.

While this plot validates the approximation for this specific time point (September 2004), it is an illustrative example. The approximation may be less precise during extreme volatility spikes where the true density could become skewed. However, this trade-off is justified, as the resulting smoothness and stability are essential for the MCMC algorithm to converge, which would be infeasible with a non-approximated particle filter.

Figure 3.4 provides this same comparison for a later time point during the high-volatility COVID-19 period (June 2020), further illustrating that even under stress, the Gaussian approximation remains reasonably robust.

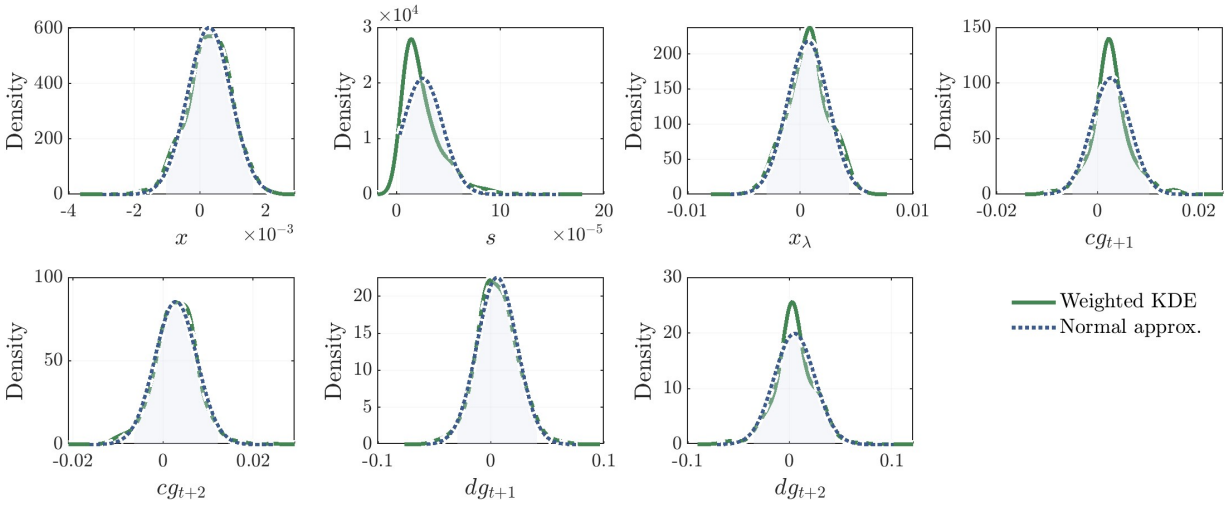



**Figure 3.3:** Comparison of the S-UPF resampling density (solid green line) and its normal approximation (dotted blue line) across seven latent dimensions during a representative stable period (September 2004).

### Likelihood Surface Smoothness and Common Random Numbers

Having established that the S-UPF yields an accurate and low-bias estimator, our final validation step addresses the stability of the inference across the parameter space. For Metropolis-Hastings MCMC to mix efficiently, the likelihood surface should not only be accurate but also sufficiently smooth with respect to the parameters. Discontinuities or high-frequency noise can cause the sampler to get stuck in local optima or reject valid moves.

To achieve this smoothness, we couple the S-UPF with CRNs. This approach fixes the random seed across likelihood evaluations, inducing positive correlation in the Monte Carlo errors and revealing the underlying geometry of the objective function.

Figure 3.5 presents the log-likelihood profiles for a selection of key parameters, holding others fixed. (Profiles for all remaining parameters are provided in Figure B.2 in Appendix B.6). The green curve represents the unbiased UPF baseline, while the blue curves represent the S-UPF-CRN at varying particle counts.

The results reveal that the S-UPF-CRN effectively preserves the global curvature of the likelihood surface, closely matching the green UPF baseline and ensuring that the maximum likelihood point remains identifiable. Crucially, this curvature remains stable and consistent with the profiles generated at higher particle counts even when using significantly

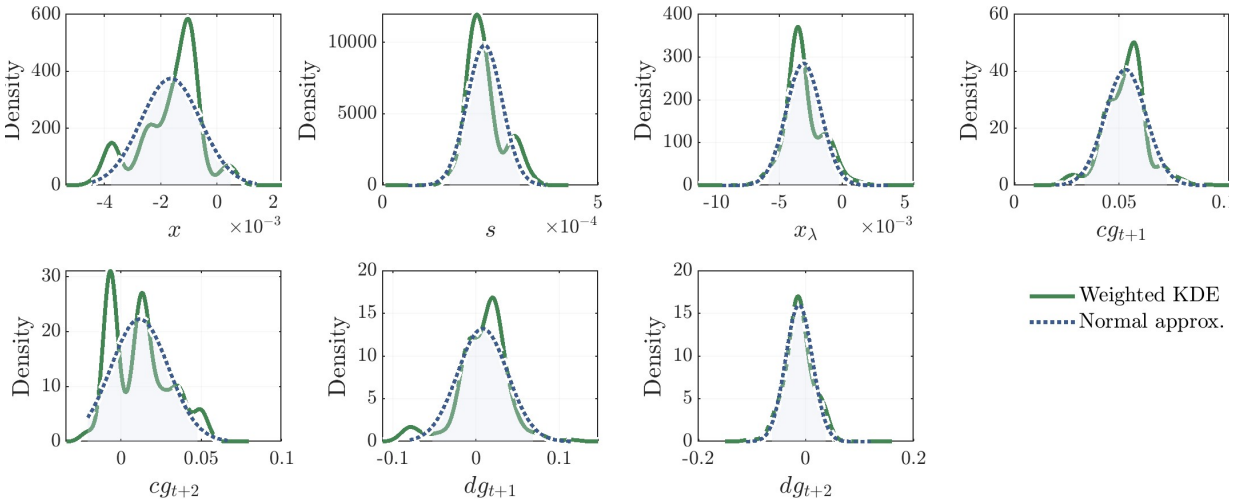



**Figure 3.4:** Comparison of the S-UPF resampling density (solid green line) and its normal approximation (dotted blue line) across seven latent dimensions during the high-volatility COVID-19 period (June 2020).

fewer particles, such as $M = 2^7$. While the Gaussian approximation and the CRN scheme naturally produce smoother surfaces by mitigating resampling discontinuities and Monte Carlo noise, the primary advantage is this preservation of the likelihood's geometric structure. This stability confirms that the S-UPF-CRN-SMC framework is well-suited for navigating the complex posterior landscapes of mixed-frequency LRR models.

### 3.4.5 SMC Configuration

The sampler, S-UPF-CRN-SMC, is configured with $2^{11}$ parameter particles and $2^9$ state particles, employing a total of 10 move steps. When bridging between the prior and posterior distributions, we adopt a Gaussian mixture proposal with a fixed number of components in the Particle Metropolis-Hastings (PMMH) steps.

The selection of the priors in Table 3.2 follows empirical calibration, ensuring proper identification of key parameters while maintaining flexibility for Bayesian inference. The preference parameters $(\delta, \gamma, \psi)$ follow truncated normal distributions, enforcing plausible economic constraints. The stochastic volatility component $(\sigma_{\bar{c}}, \phi_s, \nu)$ is calibrated to match observed consumption growth variance. Measurement error variances $(\phi_m, \phi_f)$ are set proportionally to the observed sample variance of the market return rate $Y_m$ and risk-free rate $Y_f$, scaled by constants $K_1$ and $K_2$. During posterior inference, the Gaussian mixture

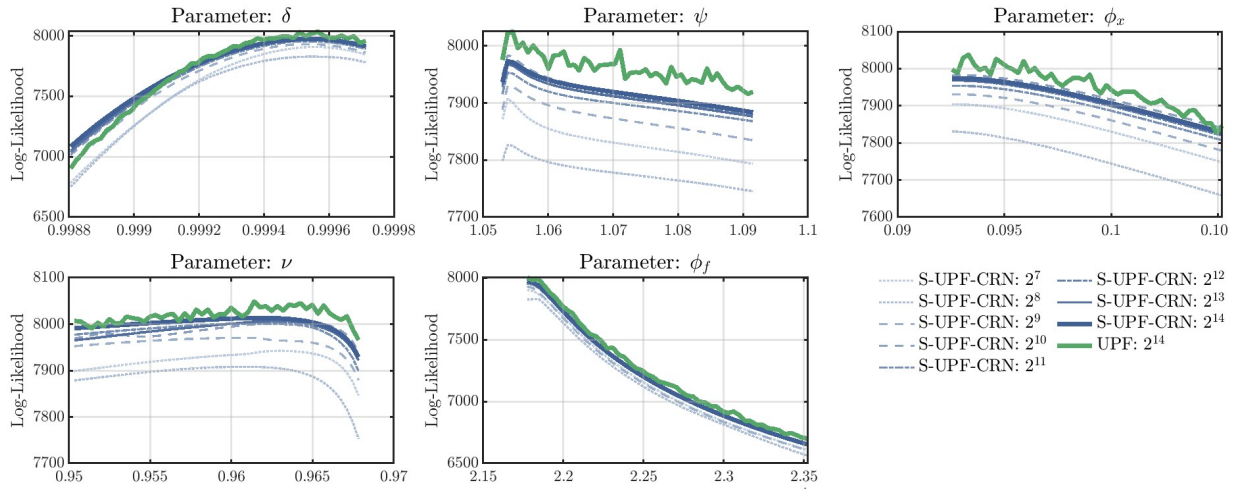



**Figure 3.5:** Log-likelihood profile plots across a selection of parameter dimensions. The benchmark UPF (green) is compared to S-UPF-CRN with different particle counts (blue curves).

**Table 3.2:** Prior Distributions for Model Parameters

| $\Theta$ | **D. Form** | **Support** | **Hyper** | $\Theta$ | **D. Form** | **Support** | **Hyper** |
|---|---|---|---|---|---|---|---|
| $\delta$ | Uniform | (0,1) | (0.80, 1.00) | $\phi_s$ | Tr. Normal | (1, $\infty$) | (2.00, 4.00) |
| $\gamma$ | Tr. Normal | (0, $\infty$) | (6.00, 2.00) | $\mu_d$ | Normal | (-$\infty$, $\infty$) | ($\bar{\mu}_d$, $10^{-8}$) |
| $\psi$ | Tr. Normal | (0, $\infty$) | (1.50, 0.50) | $\Phi$ | Tr. Normal | (0, $\infty$) | (3.00, 6.00) |
| $\rho_\lambda$ | Uniform | (0,1) | (0,1) | $\phi_{dc}$ | Tr. Normal | (0, $\infty$) | (3.00, 6.00) |
| $\sigma_\lambda$ | Tr. Normal | (0, $\infty$) | (0.002, 0.01) | $\phi_d$ | Tr. Normal | (0, $\infty$) | (5.00, 25.00) |
| $\mu$ | Normal | (-$\infty$, $\infty$) | ($\bar{\mu}$, $10^{-8}$) | $\bar{\sigma}_{sd}$ | Tr. Normal | (0, $\infty$) | (0.006, 0.005) |
| $\rho$ | Uniform | (0,1) | (0,1) | | | | |
| $\phi_x$ | Tr. Normal | (0,1) | (0.10, 0.04) | $\phi_m$ | Tr. Normal | (0, $\infty$) | ($K_1(Y_m)$, ($K_2(Y_m)$) |
| $\bar{\sigma}$ | Tr. Normal | (0, $\infty$) | (0.004, 0.005) | $\phi_f$ | Tr. Normal | (0, $\infty$) | ($K_1(Y_f)$, ($K_2(Y_f)$) |

proposals are updated iteratively.

### 3.4.6 Monte Carlo Simulation Studies

The results indicate that mixed-frequency models, estimated under our framework, yield consistently tighter posterior distributions compared to their balanced-frequency counterparts. This gain in precision is most pronounced for parameters governing high-frequency dynamics—such as volatility persistence and measurement error—which are otherwise weakly identified in aggregated data. However, this improved identification comes with a bias-variance trade-off: while variance is reduced, the smoothing approximation introduces a small estimation bias for certain structural parameters. This synthetic validation allows us

to disentangle the information gains of mixed-frequency data from the numerical properties of the filter.

To evaluate the overall performance of our complete inference pipeline, S-UPF-CRN-SMC, we conduct a synthetic data experiment using the exact SMC approach described in the previous section. The sampler is initialized with 800 time points, $2^{11}$ parameter particles and $2^9$ state particles, using the parameter priors, move steps, and Gaussian mixture proposals specified in Section 3.4.5. To ensure the robustness of the results, the metrics reported in Table 3.3 are averaged over 20 independent Monte Carlo runs. The average runtime in this simulation is approximately 85 hours. The synthetic data are generated from the long-run risk model using the true parameter values reported in Table 3.3. This design allows us to benchmark the estimation procedure under idealized conditions, free from real-world misspecification, and directly evaluate how well the inference framework recovers known structural features of the model.

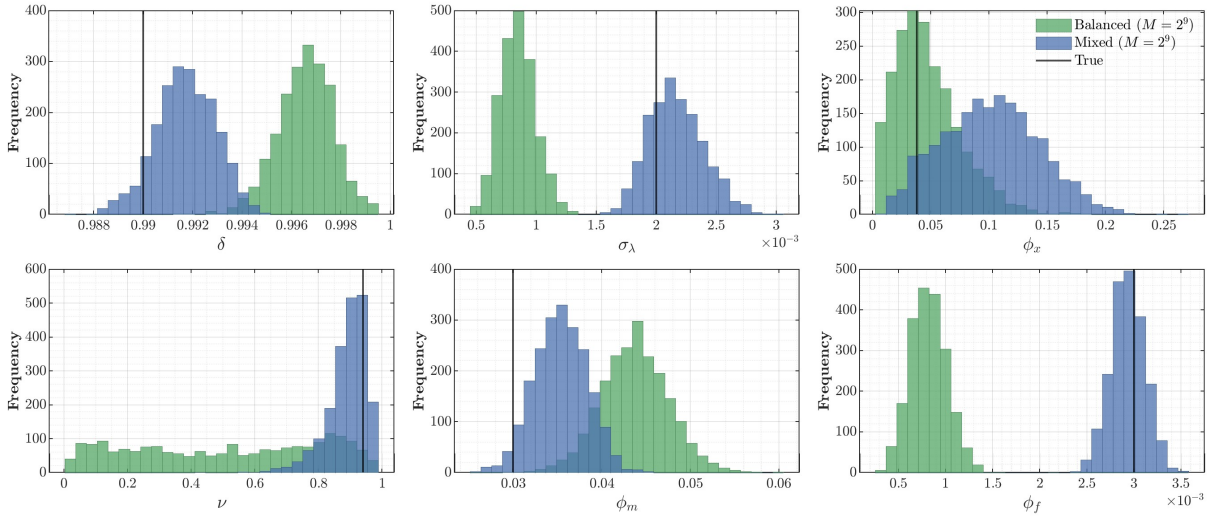



**Figure 3.6:** Posterior density estimates for six key parameters where mixed-frequency data adds significant information. Green histograms represent the balanced-frequency scenario, blue histograms represent the mixed-frequency scenario, and black vertical lines indicate the true parameter values. Note the dramatic sharpening of the posterior for $\nu$, $\phi_f$, and $\sigma_\lambda$.

The results for a representative simulation run are visualized in Figure 3.6, which displays the posterior densities for six key parameters where the impact of mixed-frequency data is most pronounced. (Posterior plots for the remaining nine parameters are provided in Figure B.3 in Appendix B.7). The quantified results are presented in Table 3.3, which reports the posterior means, standard deviations, and root mean square errors (RMSE)

alongside the true parameter values used in the data generation. To facilitate interpretation of the comparative performance, cells in the RMSE columns are highlighted in light blue where a model achieves superior estimation accuracy for a given parameter.

The visual evidence highlights the distinct advantage of the mixed-frequency model in identifying parameters that are weakly identified or biased in the balanced framework. This is most striking for the volatility and shock parameters. For instance, the balanced-frequency posterior for volatility persistence $\nu$ is diffuse and flat, suggesting that quarterly aggregated data act as a low-pass filter, smoothing out the short-term persistence of monthly volatility shocks and leaving the parameter weakly identified. Similarly, as shown in the top center panel, the balanced model systematically underestimates the true preference shock volatility $\sigma_\lambda$, likely because temporal aggregation dampens the observable volatility in the quarterly series. In contrast, the mixed-frequency posteriors for both parameters concentrate sharply around the true values, effectively solving these identification problems. Furthermore, the high-frequency data allows the model to pinpoint the measurement error variance $\phi_f$ with precision, whereas the balanced model leaves significant uncertainty.

While the mixed-frequency framework consistently reduces posterior uncertainty, our analysis reveals a discernible trade-off between variance reduction and estimation bias for specific parameters. This is most evident in the short-run growth sensitivity, $\phi_x$. As illustrated in the top-right panel, the mixed-frequency posterior is significantly tighter than the balanced posterior but exhibits a slight shift away from the true parameter value. In contrast, the balanced posterior, while centered on the truth, is highly diffuse and offers limited informative value. We attribute this minor structural bias to the Gaussian approximation within the S-UPF. The smoothing of the likelihood surface, which is essential for computational tractability, imposes a parametric assumption that simplifies the latent structure.

The increased bias in the mixed-frequency setting stems from the additional stress placed on the Gaussian approximation during state updates. In the mixed-frequency model, the filter must reconcile monthly latent dynamics with quarterly observations, effectively performing a jump across multiple time steps whenever a quarterly data point arrives. During these large updates, the highly informative nature of the high-frequency returns and interest rates can cause the true filtered density to deviate from normality. By imposing a Gaussian structure via the S-UPF to maintain a smooth likelihood surface, we simplify this latent complexity, which manifests as a minor structural bias in the posterior mean.

However, we consider this a favorable and necessary trade-off for empirical work. The substantial reduction in overall parameter uncertainty combined with the ability to identify parameters such as $\sigma_\lambda$ that are fundamentally unidentifiable in the balanced framework, provides a more robust and operationally useful basis for inference than the unbiased but prohibitively noisy alternative.

**Table 3.3:** Estimation Accuracy for a Representative Synthetic Dataset (Averaged over 20 runs)

| **Parameter** | **True** | **Balanced Frequency** | | | **Mixed Frequency** | | |
|---|---|---|---|---|---|---|---|
| | **Value** | **Mean** | **Std. Dev.** | **RMSE** | **Mean** | **Std. Dev.** | **RMSE** |
| $\delta$ | 0.9900 | 0.9966 | 0.0011 | 0.0067 | 0.9917 | 0.0012 | 0.0021 |
| $\gamma$ | 8.0000 | 7.1598 | 1.6366 | 1.8396 | 7.5168 | 1.6522 | 1.7214 |
| $\psi$ | 1.5000 | 1.1591 | 0.1608 | 0.3769 | 1.2809 | 0.1216 | 0.2506 |
| $\rho_\lambda$ | 0.9600 | 0.9393 | 0.0180 | 0.0274 | 0.9417 | 0.0121 | 0.0220 |
| $\sigma_\lambda$ | 0.0020 | 0.0008 | 0.0001 | 0.0012 | 0.0022 | 0.0002 | 0.0003 |
| $\rho$ | 0.9750 | 0.5404 | 0.2865 | 0.5205 | 0.5073 | 0.2376 | 0.5246 |
| $\phi_x$ | 0.0380 | 0.0472 | 0.0279 | 0.0294 | 0.1013 | 0.0417 | 0.0758 |
| $\bar{\sigma}$ | 0.0100 | 0.0091 | 0.0006 | 0.0011 | 0.0091 | 0.0007 | 0.0012 |
| $\nu$ | 0.9400 | 0.5101 | 0.2911 | 0.5191 | 0.8904 | 0.0620 | 0.0794 |
| $\phi_s$ | 2.0000 | 3.9350 | 1.0786 | 2.2152 | 3.3026 | 0.9466 | 1.6101 |
| $\Phi$ | 2.0000 | 4.2527 | 2.0610 | 3.0529 | 5.8346 | 2.3153 | 4.4785 |
| $\phi_{dc}$ | 2.0000 | 1.5882 | 0.4388 | 0.6016 | 1.2686 | 0.4245 | 0.8456 |
| $\phi_d$ | 5.0000 | 5.0243 | 0.4528 | 0.4533 | 5.0324 | 0.4495 | 0.4506 |
| $\phi_m$ | 0.0300 | 0.0437 | 0.0038 | 0.0142 | 0.0354 | 0.0030 | 0.0061 |
| $\phi_f$ | 0.0030 | 0.0008 | 0.0002 | 0.0022 | 0.0029 | 0.0002 | 0.0002 |

## 3.5 Data and Results

### 3.5.1 Data Description and Summary Statistics

For model estimation, we use U.S. real data from June 1959 to September 2022, covering a total of 760 months (254 quarters). We define September 2019 as the onset of the COVID-19 period, yielding a pre-COVID sample of 723 months (241 quarters) and a COVID-inclusive period spanning the final 37 months (13 quarters). We measure real consumption and dividend growth at a quarterly frequency, while risk-free rates and market returns are observed at a monthly frequency. Real consumption is constructed as the sum of seasonally adjusted real per capita nondurable and services consumption from the Bureau of Economic Analysis. Consumption growth is computed as the first difference of log-real consumption.

Real dividend growth is derived from the difference between value-weighted market returns with and without dividends. The dividend series is smoothed by aggregating four quarters' values.

The real risk-free rate is constructed from nominal 90-day U.S. Treasury bill rates and the quarterly Consumer Price Index (CPI), sourced from the Wharton Research Data Services (WRDS) Treasury and Inflation database. Ex-post real risk-free rates are obtained by deflating nominal rates using CPI inflation, and ex-ante rates are derived via regression on nominal rates and past inflation. Stock market returns are proxied by the value-weighted returns of the NYSE/AMEX/NASDAQ portfolio (VWRETD, including dividends) from the Center for Research in Security Prices (CRSP). Real stock returns are computed by deflating nominal returns with CPI inflation.

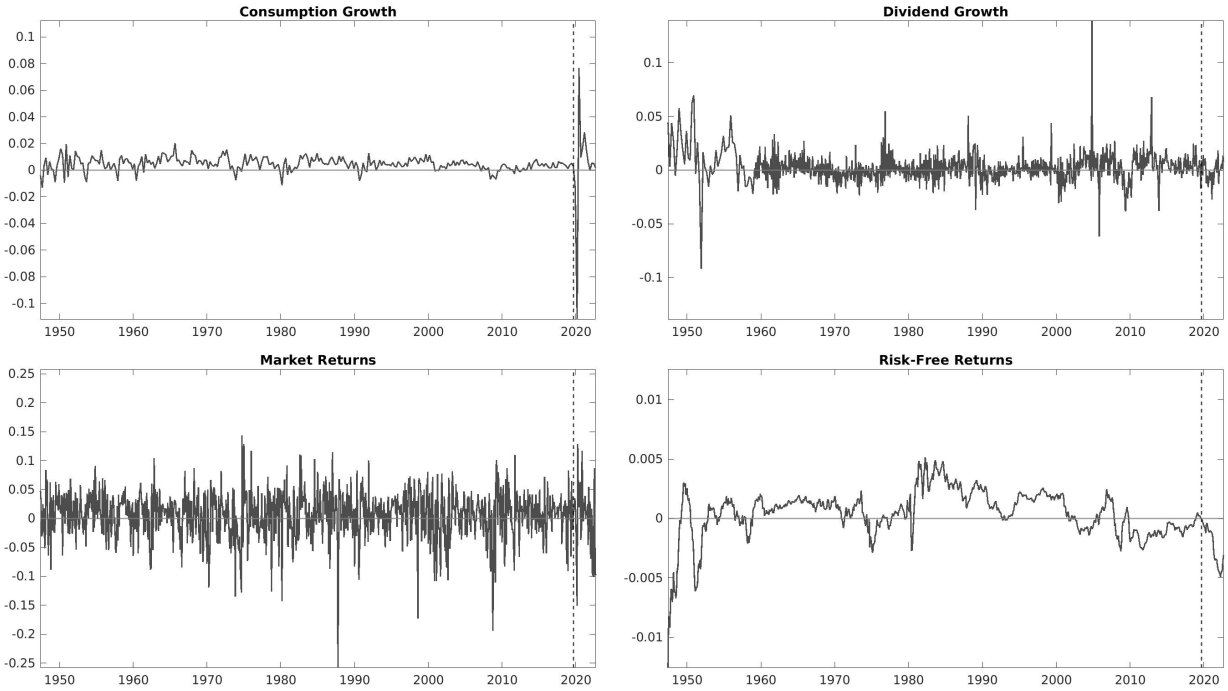



**Figure 3.7:** Time series of real consumption growth ($\Delta c$), dividend growth ($\Delta d$), real market returns ($r_m$), and real risk-free rate ($r_f$), from 1959:Q1 to 2022:Q3. The variables $\Delta c$ and $\Delta d$ are shown at quarterly frequency, while $r_m$ and $r_f$ are monthly. A vertical dashed line marks the boundary at 2019:Q3 to separate the COVID and pre-COVID-19 periods. All series are demeaned.

Table 3.4 presents summary statistics for the four main variables used in the estimation. Figure 3.7 plots the time series for these four variables, with a vertical line indicating the start of the COVID-19 period. The pre-COVID-19 period (up to 2019:09) reflects stable

economic conditions, while the COVID-19 period captures the shock-induced volatility.

**Table 3.4:** Summary Statistics

| | **Pre-COVID (up to 2019.08)** | | | **COVID-19 period (2019.09 onward)** | | |
|---|---|---|---|---|---|---|
| Statistic | Mean | Std. Dev. | ACF(1) | Mean | Std. Dev. | ACF(1) |
| $\Delta c$ | 0.0186 | 0.0099 | 0.30 | 0.0130 | 0.0817 | -0.26 |
| $\Delta d$ | 0.0095 | 0.0280 | 0.12 | -0.0017 | 0.0189 | 0.28 |
| $r_m$ | 0.0708 | 0.1480 | 0.07 | 0.0209 | 0.2086 | -0.03 |
| $r_f$ | 0.0059 | 0.0067 | 0.97 | -0.0270 | 0.0057 | 0.98 |
| $r_m - r_f$ | 0.0649 | 0.1482 | 0.07 | 0.0480 | 0.2070 | -0.05 |

Consumption growth ($\Delta c$) displays moderate persistence in the pre-COVID sample (ACF = 0.30), with a mean of 1.86% and low volatility (Std = 0.99%). During the COVID-19 period, volatility increases substantially (Std = 8.17%), while the mean declines and persistence becomes negative, consistent with the sharp economic contraction and rebound.

Dividend growth ($\Delta d$) exhibits increased volatility and higher positive skewness in the pre-COVID-19 period. During the COVID-19 shock, average dividend growth turns slightly negative, and persistence rises, suggesting that firms smoothed dividends relative to consumption.

Real market returns ($r_m$) fall sharply from an average of 7.08% to 2.09%, while their volatility increases from 14.80% to 20.86%. The equity premium ($r_m - r_f$) remains positive in both periods but shrinks in magnitude and becomes more volatile. Notably, the risk-free rate ($r_f$) turns negative during the COVID-19 period and remains highly persistent, reflecting sustained accommodative monetary policy.

The contemporaneous correlation between consumption and dividend growth is low before COVID (0.057) and turns sharply negative during the pandemic (-0.564), indicating a decoupling of the two series in response to the crisis. These patterns underscore the importance of accounting for structural breaks and support the mixed-frequency model's ability to detect and adapt to such dynamics.

### 3.5.2 Filtered States

Figure 3.8 displays the posterior means and 5–95% credible intervals for the three latent states implied by the mixed-frequency long-run risk model, while taking into account

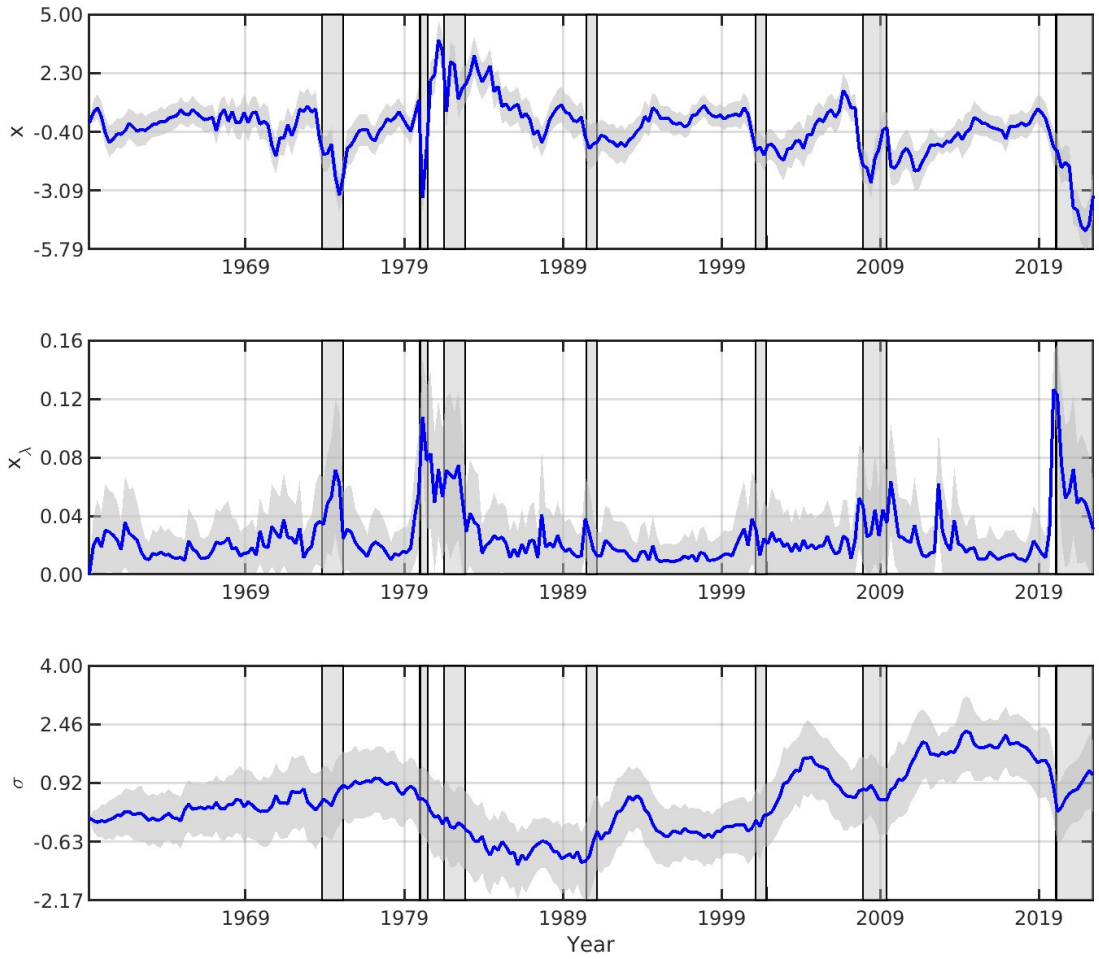



**Figure 3.8:** Filtered latent states. Posterior means (blue lines) and 5-95% credible intervals (gray shaded areas) for the latent states $x_t$, $x_{\lambda,t}$, and $\sigma_t$. Shaded vertical bars indicate NBER recessions.

parameter uncertainty. The long-run component $x_t$ moves countercyclically: it falls during each recession (1973–75, the double-dip of the early 1980s, the Global Financial Crisis, and COVID-19) and subsequently mean reverts. The within-quarter adjustment $x_{\lambda,t}$ is close to zero in tranquil periods but spikes sharply at recession onsets, with the largest spike during COVID-19, indicating rapid short-run revisions in expected growth when conditions deteriorate. The volatility state $\sigma_t$ captures the Great Moderation drop in the early 1980s, a buildup ahead of the Global Financial Crisis, and a distinct COVID-19 spike. Across all three panels the credible bands widen precisely in crisis windows and tighten in normal times, consistent with the information flow in mixed-frequency data and the inference delivered by the S-UPF-CRN-SMC.

### 3.5.3 Mixed-Frequency Estimation: COVID Impact and Structural Comparison

To assess the impact of the COVID-19 pandemic on risk perception and macroeconomic dynamics, we estimate the long-run risk model under a mixed-frequency structure using data both with and without the COVID-19 period. Table 3.5 reports the posterior means and standard deviations across the two samples, highlighting how structural parameter estimates evolve in response to the inclusion of pandemic-era data.

The inclusion of COVID-period data leads to several shifts in posterior inference. Most prominently, the risk aversion coefficient $\gamma$ declines from 4.88 to 2.68, suggesting a sharp compression in the perceived equity premium during the pandemic. This is plausibly linked to large-scale fiscal and monetary interventions, which dampened investor demand for compensation against long-run risk. Simultaneously, volatility expectations shift: the scale parameter $\bar{\sigma}$ increases while volatility persistence $\nu$ drops from 0.95 to 0.87, indicating that agents perceived the pandemic-induced volatility as transitory rather than structural. Dividend growth persistence $\phi_d$ also declines substantially, reflecting corporate payout smoothing amid earnings uncertainty.

These empirical results invite comparison with the log-linearised LRR model of Fulop et al. (2022a). Focusing on their Case 2 (Log-Linearised LRR), we observe several differences. Our no-COVID estimate of $\gamma$ (4.88) is notably lower than the 7.12 reported in Fulop et al. (2022a), reflecting both the longer post-crisis sample used here and the potential information

**Table 3.5:** Posterior Estimates from the Mixed-Frequency Model: Excluding vs. Including COVID

| | No COVID | | With COVID | |
|---|---|---|---|---|
| **Parameter** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** |
| $\delta$ | 0.99781 | 0.00069 | 0.99809 | 0.00071 |
| $\gamma$ | 4.87530 | 1.40710 | 2.68350 | 0.72970 |
| $\psi$ | 1.03270 | 0.01570 | 1.02410 | 0.01600 |
| $\rho_\lambda$ | 0.98526 | 0.00431 | 0.98922 | 0.00413 |
| $\sigma_\lambda$ | 0.000121 | 0.000015 | 0.000111 | 0.000016 |
| $\rho$ | 0.94508 | 0.01123 | 0.94343 | 0.01123 |
| $\phi_x$ | 0.09851 | 0.00583 | 0.06459 | 0.00400 |
| $\bar{\sigma}$ | 0.00308 | 0.00024 | 0.00501 | 0.00034 |
| $\nu$ | 0.94938 | 0.01448 | 0.86740 | 0.02783 |
| $\phi_s$ | 1.45130 | 0.28050 | 1.23010 | 0.11680 |
| $\Phi$ | 0.40489 | 0.19522 | 0.29782 | 0.18217 |
| $\phi_{dc}$ | 0.21512 | 0.12722 | 0.09199 | 0.06361 |
| $\phi_d$ | 2.47470 | 0.17131 | 1.54900 | 0.10480 |
| $\phi_m$ | 0.04388 | 0.00118 | 0.04462 | 0.00114 |
| $\phi_f$ | 0.000011 | 0.000008 | 0.000013 | 0.000009 |

gains from mixed-frequency estimation. Similarly, our estimated elasticity of intertemporal substitution $\psi$ is slightly below Fulop's value (1.03 vs. 1.17), implying weaker intertemporal smoothing. Our estimate of long-run growth persistence $\rho$ is 0.945, compared to 0.9805 in Fulop et al. (2022a), suggesting faster mean reversion in recent decades.

Taken together, the comparison across samples and models offers several economic insights. The COVID-19 pandemic appears as a high-impact, short-duration shock: it raised perceived volatility and depressed risk premia, yet left long-run expectations (e.g., $\rho$ and $\delta$) largely intact. Our framework captures these dynamics without conflating transitory shocks with structural parameters, validating the use of mixed-frequency models in periods of structural turbulence. For policy, this suggests that high-frequency macro-financial monitoring can improve the real-time assessment of risk and guide more timely intervention. For asset pricing, the results imply that investors reallocate portfolios based on rapidly updating perceptions of volatility, even when their long-term beliefs remain anchored.

## 3.6 Conclusion

This paper develops a robust and computationally efficient Bayesian inference framework for long-run risk models, overcoming key limitations of standard pseudo-marginal methods. By embedding a smooth unscented particle filter within an adaptive Sequential Monte Carlo sampler, and coupling it with CRNs, we eliminate the need for unbiased yet noisy likelihood estimates. The resulting S-UPF-CRN-SMC algorithm produces stable posterior inference even in the presence of nonlinear latent dynamics, highly informative observations, and mixed-frequency data.

We construct a state-space model that accommodates both quarterly and monthly observations, allowing for latent monthly processes consistent with real-world financial data structures. This mixed-frequency specification not only reflects data availability in practice but also enhances identification of short-run risk components, particularly during crisis periods. In simulation exercises using synthetic data, our framework consistently recovers the true parameters with lower estimation error and greater posterior sharpness than the balanced-frequency benchmark, confirming both accuracy and posterior concentration.

Applied to U.S. macro-financial data from 1959 to 2022, the model reveals that the COVID-19 period triggered substantial shifts in perceived risk and volatility while leaving long-run beliefs intact. We find a marked decline in the posterior mean of risk aversion, a drop in volatility persistence, higher volatility scale, and reduced dividend and short-run growth persistence during the pandemic; by contrast, the discount factor, the elasticity of intertemporal substitution, and the persistence of trend growth remain stable across samples. These findings support the interpretation of COVID as a high-impact, short-lived risk shock rather than a structural change in the economy. The filtered states corroborate this reading: the within-quarter adjustment spikes on recession onsets, the volatility factor jumps in the Global Financial Crisis and COVID episodes, and predictive bands widen precisely during those periods. In the observables, the model tracks consumption growth tightly and captures crisis-time widening in market and risk-free return uncertainty; the most extreme dividend realizations remain outside a Gaussian 90% band, pointing to heavy tails in dividend measurement.

The empirical implications extend the insights of Schorfheide et al. (2018) and Fulop et al. (2022a), while offering a more general and computationally viable framework for inference. Relative to balanced-frequency designs, mixed-frequency estimation delivers

sharper posteriors in Monte Carlo and more responsive crisis-time inference in the data, without distorting long-run components. The method captures both macroeconomic stability and asset-pricing volatility within a unified estimation structure, validating the use of mixed-frequency information even under structural breaks.

From a policy and monitoring perspective, the results highlight the importance of mixed-frequency modelling for real-time macro-financial surveillance. By disentangling transitory shocks from structural persistence, the framework provides early-warning indicators—through the short-run adjustment and volatility states—and yields predictive bands that adapt immediately to turbulence. This is especially relevant in the aftermath of crises, when traditional low-frequency models may underreact to rapid shifts in sentiment or volatility.

Looking ahead, future research could replace the log-linear solution with fully nonlinear global methods, such as projection-based solvers; incorporate heavier-tailed measurement for dividends or time-varying idiosyncratic dividend volatility to improve predictive coverage of extreme events; and implement a pre-COVID train / COVID test to further document real-time performance. Extending the framework to international panels—leveraging high-frequency data across countries—could further enhance its applicability. More broadly, the S-UPF-CRN-SMC framework opens the door to fast, scalable, and resilient macro-finance estimation under uncertainty, volatility, and mixed sampling structures.

# Chapter 4

# Credit Recovery Modelling via Sequential Monte Carlo Optimization

**Authors.** Jin-Chuan Duan, Andras Fulop, Patrick Tang.

**Abstract.**

We present a new recovery rate modelling framework via a Sequential Monte Carlo based combinatorial optimization technique, in which a small set of predictive features is selected from a large set of potential variables. This $\ell_0$-sparse selection problem is mapped into sampling using a discrete distribution over combinations of predictors, and then executed via density/probability-tempered SMC techniques. Our objective is to devise a parsimonious recovery rate model that employs a user-defined distribution, applied to a rich dataset of post-default bond recovery rates for US dollar-denominated bonds. We exploit the inherent flexibility of this optimization engine to conduct a consistent evaluation of two competing conditional specifications, the logit-transformed heteroskedastic model and the beta model, within an identical selection environment. Applied to a diverse range of up to 266 potential variables, this comparative analysis reveals a systematic discrepancy between likelihood-based metrics and predictive accuracy: commonly used log-likelihood metrics favor the logit specification, while the beta model yields superior predictive accuracy ($R^2$) and passes formal calibration tests. Furthermore, we demonstrate that incorporating firm-specific variables significantly enhances model performance. Although the algorithm can accommodate very large, collinear predictor sets, we find that light prescreening further

improves performance and interpretability. Ultimately, the analysis confirms the SMC framework as a robust and flexible method for variable selection, capable of handling complex data structures to deliver parsimonious and economically interpretable models.

## 4.1 Introduction

Understanding the factors impacting corporate default risk is crucial for managing credit risk and shaping financial regulation and macroprudential policies. This risk primarily comprises two components: the probability of default (PD) and post-default recovery rates. While the former has been extensively studied, as demonstrated by works such as Duffie et al. (2007) and Duan et al. (2012), the latter remains less explored. In this paper, we introduce a recovery modelling framework specifically designed to select an optimal subset from an extensive set of potential explanatory variables, aiming to enhance the prediction accuracy of recovery rates. We cast this selection task as an $\ell_0$-sparse model selection problem and resolve the resulting NP-hard optimization via density- and probability-tempered Sequential Monte Carlo (SMC), mapping the combinatorial optimization into sampling from a discrete target over predictor combinations. A key aim is to make it feasible to include virtually all available information—including a rich set of interaction terms—while still selecting a small, interpretable subset.

Existing recovery modelling studies, such as those by Duan and Hwang (2014), Calabrese (2014), or Bellotti and Crook (2012), often limit themselves to a constrained set of pre-selected predictors, driven by economic intuition and informed by previous research. They could potentially overlook valuable insights because there may exist a much broader list of default predictors for at least two reasons. Firstly, standard models suggest that predictors are linked to recovery through a linear index—an assumption one can naturally relax by incorporating cross products of covariates as potential predictors (interaction terms), thereby allowing for heterogeneous effects. In our application, these interaction terms turn out to be first-order: models that explicitly allow them achieve materially better predictive performance. Secondly, there are several potential proxies for each recovery driver. Rather than manually selecting from these, it may be more desirable to include all of them in the list of potential predictors and allow the algorithm to determine the optimal choice.

Several alternative strategies have been proposed in the literature to accommodate a

potentially nonlinear predictive function and/or a high-dimensional feature space. Deep learning models, as outlined by Goodfellow et al., 2016, propose a multi-layered architecture incorporating intermediate hidden units that bridge the gap between predictors and output. However, within the context of recovery modelling, these flexible nonlinear models present two significant challenges. First, given that they allow for very general nonlinear functions of the predictors, they can be hard to interpret, being more black-box type tools. Second, the inherent flexibility of these models may be problematic, since they generally need substantial training data, but the sample size available for recovery modelling is typically limited.

In this paper we adopt the approach proposed by Duan (2019), who proposes selecting an optimal subset from a plethora of potential explanatory variables in linear regressions using zero-norm penalization. He shows that the selection problem is analogous to sampling from a discrete probability function defined over all possible combinations of potential variables, with the peak representing the optimal combination, and solves it via density-tempered SMC. Here, we extend this methodology in two respects. Firstly, Duan (2019) works with linear regressions where, given a chosen set of predictors, the solution is available analytically. In contrast, recovery modelling is inherently nonlinear, as recovery rates are constrained to the unit interval; hence, given a fixed set of predictors, we resort to iterative optimization methods to obtain parameter estimates. Secondly, to mitigate parameter overfitting, we maximize a modified cross-validated objective within the SMC procedure, while retaining the same discrete-target, tempered-sampling paradigm. Practically, this include-all design is complemented by optional light prescreening to reduce redundancy and multicollinearity. While statistical performance remains robust without prescreening, we find that prescreening improves out-of-sample accuracy and, importantly, enhances interpretability by clarifying economically meaningful effects.

In this study, we explore the influence of firm-specific variables on the performance of recovery models for defaulted corporate bonds. Specifically, we are interested in understanding how the inclusion of these variables affects the predictive accuracy of our models. To this end, we focus on a comprehensive dataset that includes all interaction terms and firm-specific variables, comprising 266 predictors. Additionally, we provide a short analysis of how the inclusion of firm-specific variables affects model performance. The empirical dataset under consideration comprises 2042 defaulted US dollar-denominated corporate bonds, collected from 1998 to April 2023 by the Asian Institute of Digital Finance at

the National University of Singapore. The original dataset encompasses eight continuous and three categorical variables per entry, encapsulating details pertaining to maturity, bond type, and prevailing market conditions. Categorical variables, namely the bond's redemption type, seniority, and the issuing company's industry sector, are transformed into dummy variables for analytical purposes. After rigorous preprocessing to remove outliers and standardize the dataset, we analyze the data to uncover industry-specific default trends, with noticeable clusters in the real estate sector from 2021–2022 and the energy sector from 2015–2020. However, maturity and seniority did not exhibit a clear pattern in relation to defaults. In our empirical exercises we compare full designs that retain all predictors and interactions with prescreened designs that lightly prune redundant inputs.

We apply our SMC methodology to the dataset and evaluate both a logit-transformed heteroskedastic specification and the empirically often-used beta model. Our empirical findings yield several key insights: (a) there is a performance gap that is particularly significant in datasets enriched with firm-specific and interaction variables, underscoring the importance of accounting for heterogeneous effects in recovery modelling; (b) the robustness of our results holds across different specifications of the conditional recovery model; and (c) despite this robustness, all models exhibit some degree of misspecification, as indicated by both formal statistical tests and informal graphical diagnostics, with the specification incorporating interaction effects and firm-specific variables demonstrating the least misspecification. In our data, the beta specification attains slightly higher out-of-sample $R^2$ and passes the Conditional Kolmogorov calibration test, while the logit-heteroskedastic model delivers comparable performance; the gap between the two models is small, and both benefit from the inclusion of firm-specific and interaction terms. Finally, we show that optional prescreening improves both predictive performance and interpretability by reducing redundancy without altering the main economic conclusions.

The paper is organized as follows. In Section 4.2, we introduce a specific recovery model and emphasize the importance of zero-norm penalization. In Section 4.3, we outline our proposed methodology for variable selection based on SMC and explain the natural transformation of the zero-norm penalized optimization problem into a sampling task. Section 4.4 details the empirical analysis, describing the data preprocessing via rank transformation and the construction of four nested covariate designs. We then evaluate the predictive accuracy and calibration of the logit-heteroskedastic and beta models across these progressively enriched information sets. Finally, we summarize our findings in Section 4.5.

## 4.2 Recovery Modelling

We consider a regression model with $p$ regressors and $n$ observations. Here, $\mathbf{R} = (R_1, \ldots, R_n)$ and $\mathbf{X} = (X_1, \ldots, X_n)'$ represent the vector of observations of size $n \times 1$ and the regressor matrix of size $n \times p$, respectively. The recovery rate $R_i$ is subject to taking values between 0 and 1, and it is distributed according to a certain distribution, dependent on a regression coefficient $\theta$. We aim to construct a model that predicts the recovery rate $R_i$, given the predictors $X_i$.

At this stage, we do not specify the exact model types, as this general framework applies to both linear and nonlinear regression models. The target $g(\cdot)$ optimization could either be minimizing the regression's mean squared error or maximizing the log-likelihood or their modified versions, depending on the problem formulation. The regression model with zero-norm penalization is expressed as:

$$\begin{aligned} \arg\max_{\theta} \quad & g\big(\theta; \mathbf{R}, \mathbf{X}\big) \\ \text{s.t.} \quad & |\theta|_0 \leq p_s \leq p \end{aligned} \tag{4.1}$$

where $|\theta|_0$ counts the number of non-zero entries in $\theta$ and the optimal number of predictors is denoted by $p_s$.

Although this zero-norm penalized regression problem is known to be NP-hard, as stated by Natarajan (1995), there exist numerous reasons to favor zero-norm penalization over other forms of penalization. $\ell_1$-norm penalization has become a prevalent tool in the statistical literature for addressing issues of in-sample overfitting and facilitating variable selection. This popularity is primarily attributable to its computational advantages due to the convexity of its objective function. Our proposed framework employs zero-norm penalization, offering a more theoretically attractive solution with significant advantages over $\ell_1$ penalization. This zero-norm penalized regression is natural for regressor selection, since it preserves the integrity of regression coefficients. Unlike the $\ell_1$ approach, it doesn't cause shrinkage towards zero upon selection, hence eliminating the distortion of coefficient values by the penalty term. Finally, an intriguing aspect of zero-norm penalization is its invariance to linear transformations of a group of regressors. Unlike the $\ell_p$ penalty term (where $0 < p < 2$), which is sensitive to such transformations, the zero-norm penalty remains unaffected. This insensitivity to multicollinearity, which naturally occurs in data,

ensures that regressor selection based on the zero-norm penalty remains unchanged.

Next, we consider two specifications for the conditional distribution of recoveries: (i) a beta regression and (ii) a logit-transformed linear model with heteroskedastic errors. In both cases we use two linked parts—a mean equation and a dispersion (variance/precision) equation—and we allow covariates to enter each part. Thus, predictors can affect not only the expected recovery but also how variable recoveries are around that expectation. The SMC procedure selects variables separately for the mean and dispersion components.

### 4.2.1 Beta Regression

Beta distributions provide a straightforward and parsimonious method to capture a wide array of distributional shapes within the unit interval. Therefore, we assume the recovery rate, constrained to have values between 0 and 1, to follow a calibrated beta distribution:

$$\begin{aligned} E(R_i) &= \mu_i, \quad 0 < \mu_i < 1 \\ Var(R_i) &= \frac{\mu_i(1-\mu_i)}{\phi_i + 1}, \quad \phi_i > 0 \\ h(R_i, \mu_i, \phi_i) &= \frac{\Gamma(\phi_i)}{\Gamma(\mu_i\phi_i)\Gamma(\phi_i - \mu_i\phi_i)} R_i^{\mu_i\phi_i - 1}(1-R_i)^{\phi_i - \mu_i\phi_i - 1}, \end{aligned}$$

where the mean, transformed using logit, and the precision $\phi_i$, transformed using log, can be then parameterized as:

$$\ln \frac{\mu_i}{1-\mu_i} = \eta' X_i, \qquad \ln \phi_i = \rho' X_i.$$

It is important to distinguish this specification from standard variance modelling. Here, we model the precision parameter $\phi_i$. Since $Var(R_i) \propto 1/(1+\phi_i)$, $\phi_i$ is inversely related to the variance; thus, a positive coefficient in $\rho$ implies higher precision and lower conditional variance.

Let $\theta = \{\eta, \rho\}$ denote the unknown parameter vector; we aim to find the combination of predictors with the highest likelihood. The log-likelihood of the beta-regression is expressed

as:

$$\mathcal{L}(\theta; \mathbf{R}, \mathbf{X}) = \sum_{i:R_i \in (0,1)} \Bigg[ \ln \Gamma(e^{\rho' X_i}) - \ln \Gamma\left(\frac{e^{\eta' X_i + \rho' X_i}}{1 + e^{\eta' X_i}}\right) - \ln \Gamma\left(\frac{e^{\rho' X_i}}{1 + e^{\eta' X_i}}\right) + \left(\frac{e^{\eta' X_i + \rho' X_i}}{1 + e^{\eta' X_i}} - 1\right) \ln R_i + \left(\frac{e^{\rho' X_i}}{1 + e^{\eta' X_i}} - 1\right) \ln(1 - R_i) \Bigg].$$

Similar to Eq. (4.1), we are able to formulate a maximization problem, such that

$$\begin{aligned} \arg\max_{\theta,\rho} \quad & \mathcal{L}(\theta; \mathbf{R}, \mathbf{X}) \\ \text{s.t.} \quad & ||\theta||_0 = ||\eta||_0 + ||\rho||_0 \leq p_s \leq p. \end{aligned} \tag{4.2}$$

### 4.2.2 Logit-Transformed Linear Regression

Traditional linear regression techniques are not directly applicable due to the bounded nature of the dependent variable. To address this, a logit transformation is applied to $R$ to map it onto the real line, thereby facilitating the use of linear regression techniques. Mathematically, the logit transformation is defined as:

$$\text{logit}(R_i) = \log\left(\frac{R_i}{1 - R_i}\right)$$

This transformation is invertible, ensuring that predictions can be mapped back to the original scale of $\mathbf{R}$. Subsequent to this transformation, a linear regression model is fitted to the transformed dependent variable, logit($\mathbf{R}$), with a set of predictors encapsulated in the design matrix $\mathbf{X}$. The model can be formally expressed as,

$$\text{logit}(R_i) = X_i \beta_{OLS} + \xi_i,$$

where we assume the error term $\xi_i$ follows a normal distribution and its variance is modelled as a function of the predictors, using an exponential function to ensure positivity, such that

$$V(\xi_i | X_i) = \exp(X_i \kappa),$$

where $\kappa$ is a vector of unknown parameters to be estimated. In this specification, a positive coefficient in $\kappa$ implies higher conditional variance.

The likelihood function for this heteroskedastic model is then given by:

$$L_{GLS}(\beta_{GLS}, \kappa|\mathbf{R}, \mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \exp(X_i\kappa)}} \exp\left(-\frac{(\text{logit}(R_i) - X_i'\beta)^2}{2\exp(X_i\kappa)}\right) \times \frac{1}{R_i(1-R_i)}$$

The log-likelihood, which will be optimized, is given by:

$$\begin{aligned}\mathcal{L}_{GLS}(\beta_{GLS}, \kappa|\mathbf{R}, \mathbf{X}) = &-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log(\exp(X_i\kappa)) - \frac{1}{2}\sum_{i=1}^{n}\frac{(\text{logit}(R_i) - X_i'\beta_{GLS})^2}{\exp(X_i\kappa)} \\ &+ \sum_{i=1}^{n}\log\left(\frac{1}{R_i(1-R_i)}\right).\end{aligned} \tag{4.3}$$

Maximizing this log-likelihood with respect to $\beta_{GLS}$ and $\kappa$ yields the Maximum Likelihood Estimates of the parameters, accounting for heteroskedasticity in the model.

Note that in the homoscedastic nested model, when we assume a constant variance, this MLE estimator is equivalent to the simple OLS estimator:

$$\begin{aligned}\hat{\beta}_{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{logit}(\mathbf{R}) \\ \hat{\sigma}^2_{OLS} &= \frac{1}{n}(\text{logit}(\mathbf{R}) - \mathbf{X}\hat{\beta}_{OLS})'(\text{logit}(\mathbf{R}) - \mathbf{X}\hat{\beta}_{OLS}).\end{aligned} \tag{4.4}$$

## 4.3 Variable Selection using SMC

The SMC sampling technique that we will employ, while rooted in Bayesian literature, deviates significantly from traditional methodologies employing Bayesian statistical tools. Typically, such methods hinge on a hierarchical structure, where each regression coefficient is modelled using a spike-and-slab (i.e., Bernoulli and Gaussian) mixture or its variants, as demonstrated in seminal works by Mitchell and Beauchamp (1988), George and McCulloch (1993), and Polson and Xu (2018). Another approach involves adaptive sampling for the inclusion of variables, driven by a parametric distribution for binary variables, a technique exemplified by Schäfer and Chopin (2013). However, these conventional Bayesian algorithms often necessitate strong priors and make additional distributional assumptions.

As a consequence, they alter the original variable selection problem. In contrast, our SMC-based variable selection method refrains from making any such supplemental distributional assumptions, directly tackling the combinatorial problem in its original form.

The primary idea is to first devise a density-tempered SMC method of finding the optimal regressor combination and the associated parameter estimates under a fixed number of regressors. Subsequently, we detail the process of determining the correct number of regressors to prevent in-sample overfitting. We transform the problem discussed above into a maximization problem over a distribution function defined on a $p_s$-dimensional random vector, $\mathbf{U} = (U_1, \ldots, U_{p_s})$, taking values from the set of regressor sequence numbers $\mathbf{P} = \{1, \ldots, p\}$ without replacement. Importantly, the solution in Eq. (4.1) will clearly be at the boundary, indicating that the inequality constraint can be converted into an equality. The equivalent reformulation is:

$$\arg \max_{\mathbf{U} \in \mathbf{P}(p_s)} \quad \hat{g}\big(\hat{\theta}(\mathbf{R}, \mathbf{X}_\mathbf{U}); \mathbf{R}, \mathbf{X}_\mathbf{U}\big),$$

where $\mathbf{P}(p_s)$ denotes $\mathbf{U} \in \mathbf{P}^{p_s}$, for $U_1 \neq U_2 \neq \ldots \neq U_{p_s}$, with $\mathbf{P}^{p_s}$ representing the $p_s$ Cartesian product of the set of the $\mathbf{P}$ variables, $\mathbf{X}_\mathbf{U}$ denotes the sub-matrix of $\mathbf{X}$ whose columns correspond to the regressor sequence numbers in $\mathbf{U}$, and $\hat{\theta}(\mathbf{R}, \mathbf{X}_\mathbf{U}) = \max_\theta g(\mathbf{R}, \mathbf{X}_\mathbf{U})$ stands for the optimal regression $\theta$ when $\mathbf{U}$, $\mathbf{R}$ and $\mathbf{X}_\mathbf{U}$ are known. In this paper, we use the Broyden–Fletcher–Goldfarb–Shanno algorithm to solve such unconstrained nonlinear optimization problems.

We will define the reformulated objective function $\hat{g}(\cdot)$ more closely in the subsequent sections, as we introduce a scheme to account for the parameter uncertainty, which essentially is a cross-validation scheme embedded within the optimization. We define the cross-validated objective:

$$\hat{g}\big(\hat{\theta}(\mathbf{R}, \mathbf{X}_\mathbf{U}); \mathbf{R}, \mathbf{X}_\mathbf{U}\big) = \sum_{k=1}^{K} g\big(\hat{\theta}(\mathbf{R}^{(\tilde{k})}, \mathbf{X}_\mathbf{U}^{(\tilde{k})}); \mathbf{R}^{(k)}, \mathbf{X}_\mathbf{U}^{(k)}\big).$$

For a particular $\mathbf{U}$, we divide $\mathbf{X}_\mathbf{U}$ into $K$-folds following a certain rule and evaluate $g(\cdot)$ on the current one fold, $\mathbf{X}_\mathbf{U}^{(k)}$, using the optimal coefficients $\hat{\theta}(\mathbf{R}^{(\tilde{k})}, \mathbf{X}_\mathbf{U}^{(\tilde{k})})$ trained on the remaining $K-1$ folds, denoted by $\mathbf{X}_\mathbf{U}^{(\tilde{k})}$. The sum of this $K$-fold cross-validation becomes the new objective. We aim to find a subset of predictors with their corresponding parameters that are stable across different sections of the data, thereby avoiding overfitting.

Given that $\hat{g}(\cdot)$ is positive, it can be perceived as a discrete distribution function of $\mathbf{U}$, up to a normalizing constant, over $\mathbf{P}(p_s)$, i.e., the set of permutations. As the order of regressors for a given combination yields the same regression solution, the distribution defined over permutations is a proportionally scaled down version of the distribution function defined over combinations where the scaling factor is $p_s!$. As permutations are easier to generate, we will target the distribution function over permutations, where the SMC algorithm is intrinsically independent of any proportionality constant. Our target discrete distribution function is:

$$f(\mathbf{U} \in \mathbf{P}(p_s)); \mathbf{R}, \mathbf{X}) \propto \hat{g}(\hat{\theta}(\mathbf{R}, \mathbf{X_U}); \mathbf{R}, \mathbf{X_U}). \tag{4.5}$$

The task of selecting regressors has now been transformed into finding the maximum of the above distribution function. The fundamental concept of finding the maximum of Eq. (4.5) is to generate a sample that appropriately represents this distribution function. Our distribution-tempered (SMC) approach for finding this maximum is derived from Del Moral et al. (2006) and Duan and Fulop (2015). Our strategy bears more resemblance to that of Duan and Zhang (2016). The detailed descriptions of the SMC algorithm can be found in Appendix C.1.

In our current study, the task is to find the maximum value of $f(\mathbf{U} \in \mathbf{P}(p_s); \mathbf{R}, \mathbf{X}))$ by sequentially generating $\mathbf{U}$ by tempering the distribution to arrive at a sample that accurately represents $f(\mathbf{U} \in \mathbf{P}(p_s); \mathbf{R}, \mathbf{X}))$. Hence, its maximum value and corresponding maximizer, up to a Monte Carlo error, become readily available. The described methodology extends the concept of regressor selection with penalization by transforming it into a distribution optimization problem. Leveraging the SMC approach, we circumvent the issues of high dimensionality and correlation among the predictors. Furthermore, by embedding a cross-validation scheme within the optimization process, we account for parameter uncertainty and work towards avoiding overfitting. The reformulated problem also allows us to exploit the benefits of generating permutations rather than combinations, simplifying the process without losing the essence of the problem.

In Appendix C.1, we describe a density-tempered SMC used to find the optimal regressor combination and the associated parameter estimates under a fixed number of regressors. The algorithm implemented here aligns with the one described in Duan (2019). In practice, one still needs to determine the right number of regressors to avoid in-sample overfitting

due to deploying too many variables. To do so, we start from a small $p$ and increase the steps by one each time. Meanwhile, we also track the standard error and the corresponding $t$-statistic of the parameter coefficients. Once we observe that at least one of the selected predictors is statistically insignificant, we halt the process as this latter suggests that the current set of predictors contains a superfluous one. Then, the chosen $p^*$ is the one in the previous step.

## 4.4 Empirical Application

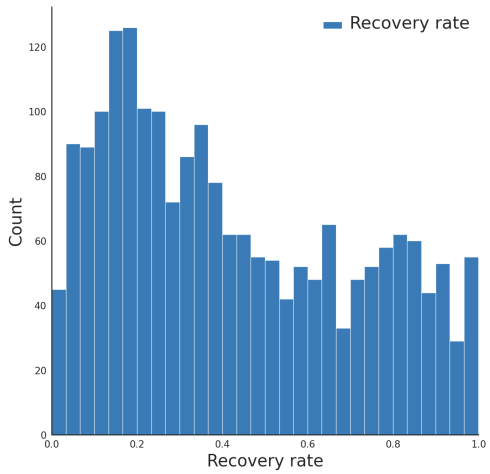
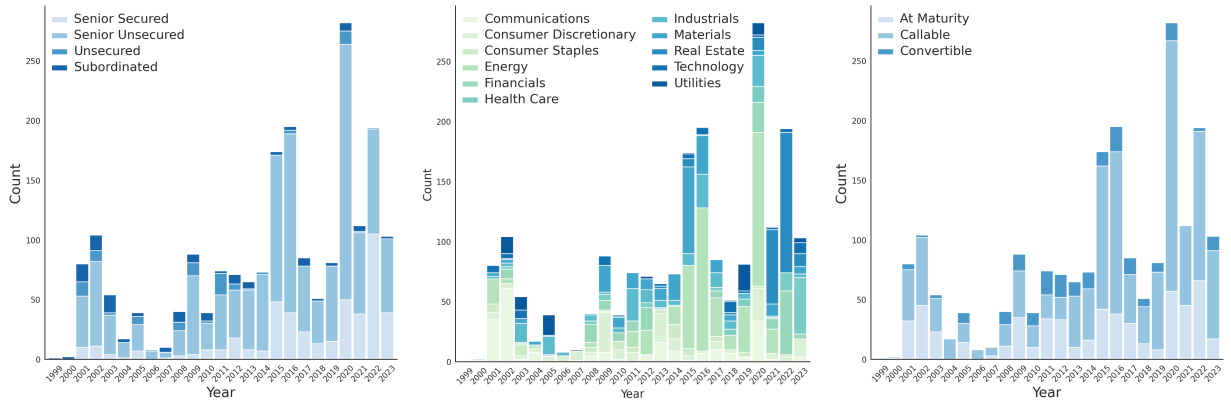



**Figure 4.1:** Histogram of recovery rates for 2,042 USD-denominated corporate bonds defaulted between 1998 and April 2023.

Our dataset comprises 2042 bonds of defaulted US dollar denominated corporate bonds, starting from 1998 to April 2023, collected by the Asian Institute of Digital Finance at National University of Singapore. Recovery values are established by dividing the post-default market values of the bond by its face value. In particular, the recovery rate is calculated as the average ratio of the bond's dirty price to its adjusted par value, assessed over one week post-default. We utilize dirty prices to capture the full economic recovery value available to the bondholder at the moment of default, inclusive of any accrued interest claims. While clean prices are standard for comparative yield analysis, the dirty price provides the most accurate measure of realized economic loss. As a consequence, as shown in Figure 4.1, we do not observe a clustering of observations at the distributional boundaries of zero recovery and full recovery, as would be the case if recovery would be defined with

ultimate recoveries.

We classify our covariates into three broad categories. First, the standard continuous predictors include: the number of days from default to maturity (*Days to Maturity*); the level of the S&P 500 index at default (*S&P500*); three U.S. Treasury interest rate series—*Interest Rate 3-Month*, *Interest Rate 1-Year*, and *Interest Rate 10-Year*—which measure short- to long-term government bond yields at the time of default; three global probability-of-default indicators—*Global PD 1-Month*, *Global PD 3-Month*, and *Global PD 12-Month*—reflecting 1-, 3-, and 12-month forward-looking default risk based on global industrial aggregates; three corresponding domestic PD measures—*Domestic PD 1-Month*, *Domestic PD 3-Month*, and *Domestic PD 12-Month*—which represent the same horizons but for issuers domiciled in the United States; and three corporate bond yield spreads from Bloomberg—*Yield Spread 3–5 Year*, *Yield Spread 5–7 Year*, and *Yield Spread 7–10 Year*—which capture credit risk premia across different maturity segments of the U.S. corporate bond market.




**(a)** Grouped by seniority **(b)** Grouped by industry **(c)** Grouped by maturity

**Figure 4.2:** Data distribution across different years.

Second, the firm-specific continuous predictors include: four historical default probability measures—*Prior PD 1-Week*, *Prior PD 2-Week*, *Prior PD 1-Month*, and *Prior PD 3-Month*—which denote the issuer's 3-month PD as observed one week, two weeks, one month, and three months prior to default; *Distance to Default 1-Day*, which refers to the issuer's distance to default one day before default, calculated from a Merton-style structural model using equity and balance sheet information from the past year; *Leverage Ratio 1-Day*, which quantifies the issuer's leverage ratio—defined as total liabilities divided by total assets—one day before default, using data from the prior year; and *Liquidity Ratio 1-Day*, which measures the issuer's liquidity position, defined as the logarithm of cash plus short-term

investments over total assets for financial firms, and the logarithm of current assets over current liabilities for non-financial firms, using the last available data within one year before the default date as a proxy.

We convert the categorical variables into dummy variables. To avoid sparsity, we merge rare subcategories and discard any dummy columns with fewer than 50 observations—most notably within the bond's redemption type and seniority. The final coding yields dummies for bond redemption type (*At Maturity*, *Callable*, *Convertible*), seniority structure (*Senior Secured*, *Senior Unsecured*, *Unsecured*, *Subordinated*), and the issuer's industry sector (*Communications*, *Consumer Discretionary*, *Consumer Staples*, *Energy*, *Financials*, *Health Care*, *Industrials*, *Materials*, *Real Estate*, *Technology*, *Utilities*). To facilitate interpretation, the seniority dummies are grouped into economically meaningful categories that reflect differences in creditor priority and claim structure.

We observe that defaults tend to occur in waves, as illustrated in Figure 4.2, which displays the distribution of bond defaults across years, disaggregated by seniority, industry sector, and maturity type. The temporal structure of the data reveals that defaults are not uniformly distributed over time but instead exhibit distinct clustering behaviour.

Panel (a) shows that bonds with *Senior Unsecured* status dominate the sample, particularly during pronounced default episodes such as those occurring in 2008–2009, 2016–2019, and again around 2020–2022. *Senior Secured* and *Subordinated* bonds appear more consistently over time, though in smaller proportions. Panel (b) further disaggregates defaults by industry. Clear sectoral waves are visible: defaults in the *Energy* sector peak between 2015 and 2020, reflecting the impact of the oil price collapse and broader commodity distress, while the *Real Estate* sector exhibits a concentrated spike in 2021–2022, possibly due to pandemic-related disruptions. The *Financials*, *Industrials*, and *Utilities* sectors also show elevated but temporally dispersed patterns, indicating more idiosyncratic distress events.

Panel (c) splits the data by bond maturity features. Here, *Callable* bonds represent the largest share of defaults across most years, especially during periods of heightened market stress, suggesting that callable structures are more common among riskier issuers or more prone to restructuring. *At Maturity* bonds maintain a steady presence, while *Convertible* bonds appear in smaller and more sporadic clusters.

While clear default waves are visible at the industry level, we do not observe strong cyclical patterns associated with bond maturity type or seniority classification. These

findings underscore the importance of industry-specific effects and structural bond features in understanding the dynamics of recovery outcomes and inform the inclusion of categorical predictors in the modelling framework.

### 4.4.1 Data Preparation

Our methodological framework is designed to handle large, collinear predictor sets. We demonstrate its robustness on the full unfiltered dataset in Appendix C.2, which includes 418 predictors (21 continuous, 18 dummies, and 378 interactions). For the main analysis, to enhance economic interpretability, we adopt a structured prescreening procedure. This primary specification, which still contains 266 predictors, is complex enough to demonstrate the model's power while facilitating a clearer discussion of economic drivers. We find this prescreened approach improves out-of-sample accuracy and interpretability compared to the full set.

#### Prescreening and Grouping

We adopt a structured data-trimming procedure prior to SMC-based variable selection, aimed at improving model interpretability and mitigating multicollinearity among strongly correlated predictors. We partition the market-level continuous covariates into four economically coherent groups—interest rates, global PD measures, domestic PD measures and U.S. corporate bond yield spreads.

Consequently, we calculate the absolute Pearson correlation between each predictor and the bond's recovery rate. Within each group, we retain only the predictor exhibiting the highest correlation magnitude. In addition, we include *Days to Maturity* throughout as a bond-level control, yielding six non-firm-specific continuous predictors for subsequent designs.

The stringent correlation-based filtering applied above is not extended to the firm-specific Prior PD horizons. We retain all these distinct issuer-level metrics because, while correlated, they collectively capture the crucial temporal evolution of issuer health leading up to default. Retaining the full set of firm-level trajectory measures is essential to maximize the local predictive power of the model, consistent with the objective of maximizing information content in the saturated designs.

### Top-ranked Variables

Within the interest rate group, *Interest Rate 1-Year* exhibits the highest univariate correlation with the recovery rate (corr = 0.044) and is therefore retained. Among the global PD measures, *Global PD 1-Month* shows the strongest correlation (corr = 0.198). Within the domestic PD group, the 12-month horizon predictor *Domestic PD 12-Month* is the most informative (corr = 0.240). In the group of U.S. corporate bond yield spreads, *Yield Spread 7–10 Year* achieves the highest correlation (corr = 0.255) and is selected for inclusion. The top-ranked variable from each group (together with *Days to Maturity*) is retained for use in all subsequent model specifications. For brevity in later paragraphs, we refer to these top-ranked variables using shortened labels that omit horizon qualifiers—specifically, *Interest Rate 1-Year* → *Interest Rate*, *Global PD 1-Month* → *Global PD*, *Domestic PD 12-Month* → *Domestic PD* and *Yield Spread 7–10 Year* → *Yield Spread*.

### Specification Grid

This trimming is applied to both the baseline non-firm-specific dataset and the augmented dataset that includes firm-specific predictors. The interaction datasets are then filtered to retain only those continuous×categorical terms whose main-effect components survived the univariate screening. These specifications are summarized in Table 4.1.

**Table 4.1:** Summary of Covariate Design Specifications.

| **Specification Name** | **Total Predictors** |
|---|---|
| *Basic* | 25 (24 regressors + 1 intercept) |
| *Basic + Inter. (NF)* | 133 (132 regressors + 1 intercept) |
| *Basic + Firm* | 32 (31 regressors + 1 intercept) |
| *Basic + Firm + Inter. (All)*[1] | 266 (265 regressors + 1 intercept) |

To systematically isolate the marginal contribution of different information sets, we define the four specifications as follows:

- *Basic:* The foundational model containing only the top-ranked market-level predictors

[1]"Interactions (Non-Firm)" in specification 2 includes only the $6 \times 18 = 108$ interactions formed from the six non-firm continuous predictors, whereas "Interactions (All)" in specification 4 includes the $13 \times 18 = 234$ interactions formed from all thirteen continuous predictors.

(e.g., *Interest Rate, Global PD*), bond-level controls (*Days to Maturity*), and the full set of categorical dummy variables.

- *Basic+ Inter. (NF):* Augments the Basic set with interaction terms derived strictly from the non-firm (NF) continuous variables multiplied by the categorical dummies.
- *Basic + Firm:* Adds the specific issuer health metrics— *Leverage Ratio, Liquidity Ratio, Distance to Default, and Prior PDs*—to the Basic design, without including any interaction terms.
- *Basic + Firm + Inter. (All):* The saturated specification. It includes all market and firm-specific predictors, along with the comprehensive set of interactions between all continuous variables and the categorical dummies.

We restrict interactions to continuous×categorical pairings for two primary reasons. First, this choice preserves interpretability: each interaction term transparently quantifies how the effect of a continuous bond characteristic, such as *Days to Maturity*, varies across different redemption types, seniority levels, or industry sectors. Second, it maintains parsimony by avoiding the rapid proliferation and collinearity issues associated with continuous×continuous interactions. For clarity, interaction terms are denoted using an asterisk to connect the interacting components—for example, *Global PD * Real Estate*. This tiered design allows us to isolate and compare the incremental explanatory power contributed by firm-specific health metrics and their interactions with bond-level features within a unified, parsimonious modelling framework.

**Rank transform**

Prior to inputting the data into the SMC, we preprocess it by applying a rank transformation to all continuous variables—a standard procedure in machine learning. This process involves converting the data into ranks and normalizing these ranks to fall between 0 and 1. The rank transformation preserves the ordinal relationships among values while reducing the impact of outliers. This is crucial when incorporating interaction terms, where continuous variables are multiplied by categorical variables, potentially introducing outliers even if the original data is free from extremes. Such outliers in the test dataset could significantly affect out-of-sample performance. By applying the rank transformation, we mitigate this risk, ensuring model robustness and enhancing prediction reliability.

### 4.4.2 Predictive Performance

For this study, we examine our proposed framework on the regression problem formulated in Eq. (4.1). We adopt a cross-sectional training and testing split, stratified by default year, such that 70% of the data from each year forms the training set, while the remaining 30% serves as the test set.

To facilitate a coherent discussion connecting predictive performance (Table 4.2) to calibration diagnostics (Figure 4.3) and economic interpretation (Section 4.4.4), we report results for a single representative canonical split. This focus is methodologically supported by our data preprocessing and sampling design. First, the stratification by year ensures that both training and testing sets remain structurally representative of the entire default timeline, preventing temporal bias. Second, as detailed in Section 4.4.1, all continuous predictors undergo a rank transformation prior to analysis. This transformation maps potentially volatile raw data into a bounded uniform domain, effectively neutralizing the leverage of extreme outliers that would otherwise induce significant variance in parameter estimates across different random partitions.

**Table 4.2:** Comparison of in-sample (IS) and out-of-sample (OOS) predictive power. Note: $R^2$ metrics are averaged over 20 random splits to ensure robustness. Log-likelihood comparisons are provided in Appendix C.2 for completeness.

| Model | Metric | Basic | Basic+Inter.(NF) | Basic+Firm | Basic+Firm+Inter.(All) |
|---|---|---|---|---|---|
| Logit-Hetero. | IS $R^2$ | 0.1833 | 0.2651 | 0.3423 | 0.4125 |
| | OOS $R^2$ | 0.1749 | 0.2364 | 0.3113 | 0.3542 |
| Beta | IS $R^2$ | 0.2105 | 0.2668 | 0.3579 | 0.3912 |
| | OOS $R^2$ | 0.2187 | 0.2241 | 0.3392 | 0.3766 |

Several key patterns emerge from these aggregate results. First, both models demonstrate improved fit with model complexity. Out-of-sample $R^2$ increases steadily across the design sequence—*Basic*, *Basic+Inter.(NF)*, *Basic+Firm*, *Basic+Firm+Inter.(All)*—highlighting the incremental explanatory power added by firm-level information and interaction terms. Notably, the logit-heteroskedastic model peaks at an out-of-sample $R^2$ of 0.3542 in the most saturated design, while the beta model peaks higher at 0.3766 in the same configuration.

Second, the beta model generally produces superior out-of-sample $R^2$ in covariate designs that include firm-specific covariates. For instance, in the *Basic+Firm* setting, the beta model attains an out-of-sample $R^2$ of 0.3392, compared to 0.3113 for the logit-heteroskedastic

model. This pattern suggests that the beta distribution's flexible functional form is better able to accommodate the complex, nonlinear variation introduced by firm-level heterogeneity. Conversely, in the simpler *Basic + Inter.(NF)* configuration, the logit-heteroskedastic model slightly outperforms in terms of out-of-sample $R^2$ (0.2364 vs. 0.2241), indicating a potential advantage in capturing sector-level heterogeneity through discrete interactions. While out-of-sample log-likelihoods were also computed, they are highly sensitive to boundary observations in the recovery distribution; for a complete technical comparison, these metrics are provided in Appendix C.3.

In sum, both models benefit from richer feature sets, but their comparative strengths differ across covariate designs. The logit-heteroskedastic model performs well in interaction-driven settings, while the beta model shows stronger results when firm-specific predictors are included. The two models thus offer complementary insights into the recovery rate structure.

### 4.4.3 Conditional Calibration and Distributional Diagnostics

In this section, we move beyond aggregate performance metrics to examine the distributional validity of the models.

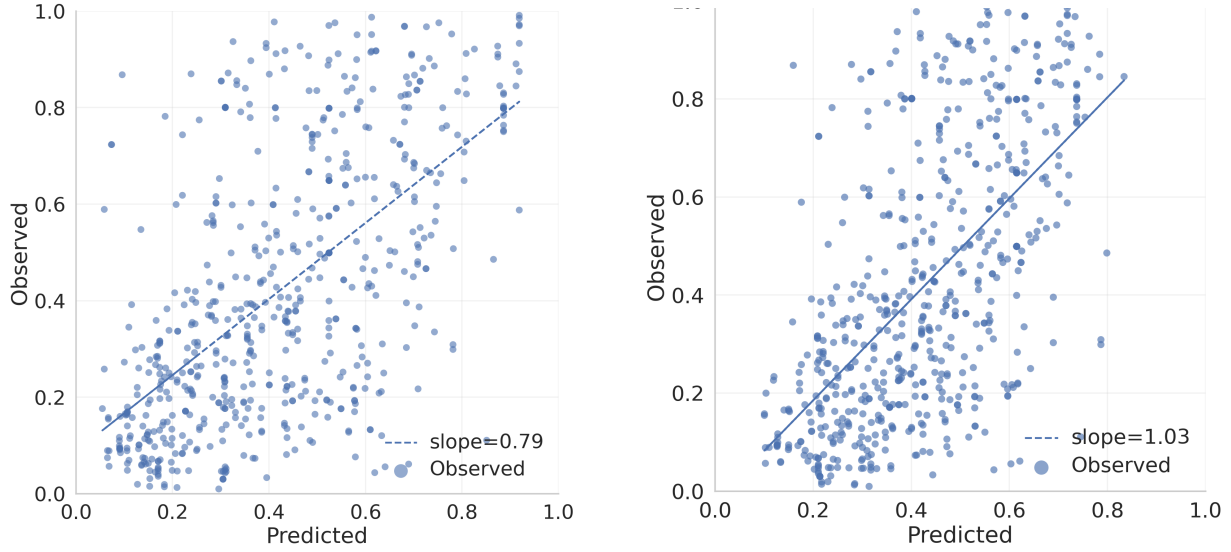
**Table 4.3:** Comparison of calibration statistics for logit-heteroskedastic and beta models. Statistics are computed on the representative training split. The null hypothesis is that the model is correctly specified.

| Model | Metric | Basic | Basic+Inter.(NF) | Basic+Firm | Basic+Firm+Inter.(All) |
|---|---|---|---|---|---|
| Logit-Hetero. | *CK statistic* | 0.5801 | 0.8083 | 0.5793 | 0.5576 |
| | *critical value* | 0.4589 | 0.6354 | 0.3806 | 0.5486 |
| Beta | *CK statistic* | 0.9042 | 1.2468 | 1.2159 | 0.3531 |
| | *critical value* | 1.5329 | 2.1376 | 2.4486 | 1.4944 |

The analysis of predictive performance and model specification, detailed in Table 4.2 and Table 4.3, reveals a notable divergence between key evaluation metrics. We examine this through formal specification testing. The beta model is the only specification for which the bootstrap-based Conditional Kolmogorov (CK) test (Andrews, 1997) does not reject the null hypothesis of correct specification. This result provides statistical support for the validity of the beta distribution in modelling the conditional recovery rates within our sample. We acknowledge, however, that these empirical findings are data-dependent; failing

to reject the null hypothesis implies that the model is consistent with the data, not that it is the true data-generating process. In contrast, the logit-heteroskedastic model is rejected by the CK test at the 5% significance level.

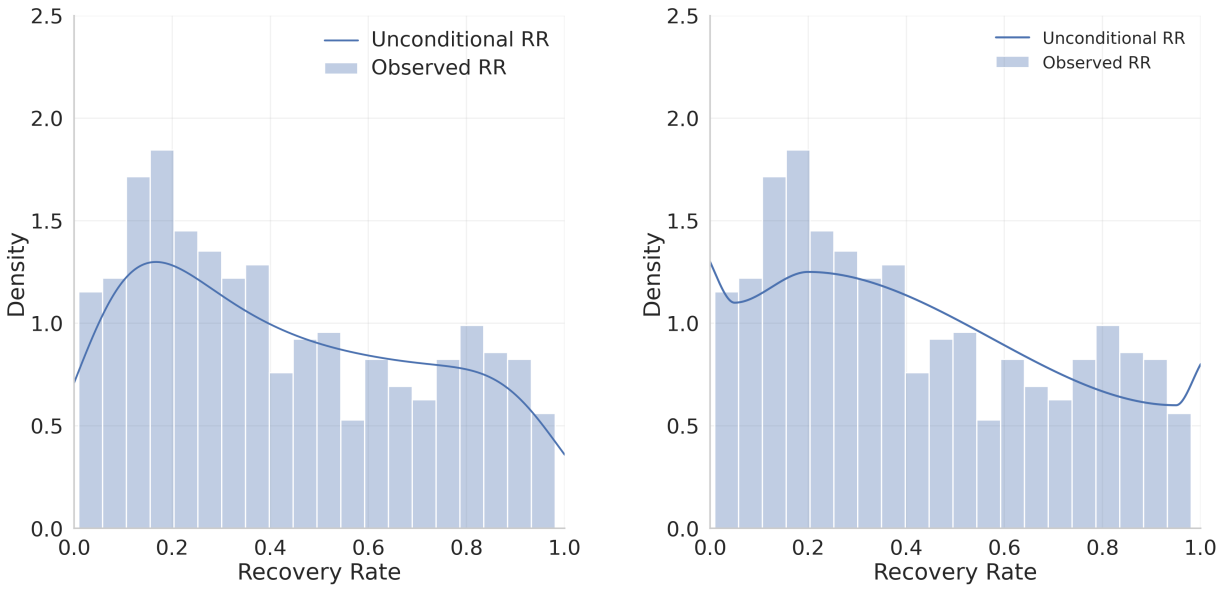
The rejection of the logit-heteroskedastic model by the CK test suggests that while it may fit certain segments of the distribution, it fails to capture the global conditional structure as effectively as the beta model. This is further reflected in the $R^2$ metrics; the beta model's superior predictive accuracy on the original economic scale confirms its suitability for characterizing the conditional mean.




**(a)** Logit-Heteroskedastic Model (Slope = 0.79) **(b)** Beta Model (Slope = 1.03)

**Figure 4.3:** Predicted vs. Observed calibration plots under the *Basic+Firm+Inter.(All)* design. The 45-degree line represents perfect calibration. Panel (a): logit-heteroskedastic model. Panel (b): beta model. Coefficient estimates are in Table 4.4 and 4.5.

Figure 4.3 presents complementary graphical diagnostics for the canonical split under the saturated design. In the predicted vs. observed panels, the point clouds for both models generally align with the 45-degree line. However, a calibration regression of observed values on predicted values reveals distinct differences: the logit-heteroskedastic model yields a slope of 0.79, while the beta model yields a slope of 1.03. The beta slope's proximity to one suggests better conditional calibration for this split. In contrast, the logit-heteroskedastic slope of less than one points to potential overconfidence, where the model generates predictions that are more extreme than the realized outcomes. This behaviour aligns with the CK test results discussed previously. While the logit-heteroskedastic model appears to




**(a)** Logit-Heteroskedastic Model **(b)** Beta Model

**Figure 4.4:** Unconditional distribution diagnostics under the *Basic+Firm+Inter.(All)* design. Compares the empirical data distribution (blue) with the model-implied unconditional distribution (orange). Panel (a): logit-heteroskedastic model. Panel (b): beta model.

capture extreme low and high recovery rates with greater sharpness, the regression slope implies that this precision at the extremes may reduce overall calibration accuracy.

Finally, Figure 4.4 compares the empirical distribution of recoveries with the model-implied unconditional distributions (obtained by averaging fitted conditional distributions over the covariates). Taken together with the CK results and the calibration slopes, these plots confirm that the beta model achieves stronger conditional calibration overall, while the logit-heteroskedastic specification offers a crisper representation of observed multimodality, albeit with a systematic deviation from the conditional mean.

### 4.4.4 Economic Implications

The SMC-selected sets under the saturated design (Tables 4.4 and 4.5) are interaction-heavy, and the two models agree on several core drivers. In both, *Days to Maturity* loads strongly and positively in the mean equation, indicating that longer-dated issues tend to realize higher expected recoveries—consistent with greater restructuring optionality and going-concern value. *Subordinated* is negative, reaffirming that junior claims recover less

**Table 4.4:** Parameter estimates (SMC-selected) for Logit-Heteroskedastic model under the *Basic+Firm+Inter.(All)* design.

| Variable | Estimate | Std. Err. |
|---|---|---|
| **Mean Equation** | | |
| *Days to Maturity* | 2.5645 | (0.1014) |
| *Domestic PD* | -1.0547 | (0.1219) |
| *Domestic PD * Financials* | 1.9898 | (0.2584) |
| *Domestic PD * Real Estate* | -0.8101 | (0.2113) |
| *Global PD * Utilities* | 2.5109 | (0.3751) |
| *Global PD * Communications* | 0.9813 | (0.1471) |
| *Interest Rate * Financials* | -1.4240 | (0.2519) |
| *Liquidity Ratio * Real Estate* | -8.5595 | (0.7626) |
| *Liquidity Ratio * Senior Unsecured* | -1.1447 | (0.1034) |
| *Subordinated* | -1.1489 | (0.1397) |
| *Yield Spread* | -1.2250 | (0.1068) |
| **Variance Equation** | | |
| *Days to Maturity * Callable* | -0.8699 | (0.1588) |
| *Global PD * Senior Unsecured* | -0.6318 | (0.1650) |
| *Global PD * Health Care* | 3.7593 | (1.3205) |
| *Liquidity Ratio * Real Estate* | -14.8373 | (1.2923) |
| *S&P500 * Unsecured* | 4.5916 | (1.1334) |
| *Intercept* | 1.1427 | (0.1242) |

than senior classes. Market stress is also visible: *Yield Spread* is negative in both models' means, so wider corporate spreads (tighter credit conditions) coincide with lower expected recoveries.

Default-risk indicators and sector heterogeneity show up through interactions. *Domestic PD* has a negative mean effect and is especially punitive for *Real Estate* (e.g., *Domestic PD * Real Estate* is negative), consistent with that sector's cyclicality and collateral illiquidity. By contrast, *Global PD * Utilities* is positive in the mean across both models, suggesting utilities are relatively insulated from global credit shocks—plausibly due to regulation and the essential nature of services. In the beta model, *Prior PD * Real Estate* is negative, indicating that elevated near-term default risk before default foreshadows weaker recoveries in real estate.

Firm health near default matters as well. Interactions with the *Liquidity Ratio* are negative in the mean (e.g., *Liquidity Ratio * Real Estate*, *Liquidity Ratio * Senior Unsecured*), implying that thin liquidity one day prior to default is associated with lower expected recoveries—and that this penalty is stronger in illiquid-collateral sectors and for unsecured

claims. Even within senior unsecured, these interactions show meaningful dispersion in outcomes once firm condition is taken into account.

**Table 4.5:** Parameter estimates (SMC-selected) for Beta model under the *Basic+Firm+Inter.(All)* design.

| Variable | Estimate | Std. Err. |
|---|---|---|
| **Mean Equation** | | |
| *Days to Maturity* | 1.9336 | (0.1066) |
| *Days to Maturity * Energy* | 1.0615 | (0.1938) |
| *Domestic PD * Energy* | -1.1881 | (0.1406) |
| *Domestic PD * Real Estate* | -1.7036 | (0.2727) |
| *Global PD * Utilities* | 2.7427 | (0.4431) |
| *Prior PD 2-Week * Real Estate* | -1.4971 | (0.3665) |
| *Liquidity Ratio * Senior Unsecured* | -1.2259 | (0.0938) |
| *Subordinated* | -1.2877 | (0.1473) |
| *Yield Spread* | -1.2306 | (0.0840) |
| **Dispersion (Precision) Equation** | | |
| *Days to Maturity * Energy* | -1.5287 | (0.1766) |
| *Global PD * Real Estate* | -4.3377 | (0.3783) |
| *Liquidity Ratio * Senior Unsecured* | -0.6954 | (0.1795) |
| *S&P500* | 1.0424 | (0.1863) |
| *Intercept* | 0.7640 | (0.1328) |

The dispersion equations add nuance about predictability. In the logit-heteroskedastic model, *Days to Maturity * Callable* and *Global PD * Senior Unsecured* are negative in the dispersion equation, consistent with lower conditional dispersion (more tightly clustered outcomes) under those conditions, whereas *Global PD * Health Care* and *S&P500 * Unsecured* are positive, consistent with higher dispersion (more heterogeneous outcomes) for those groups. A particularly large negative dispersion coefficient is *Liquidity Ratio * Real Estate*, suggesting that when near-default liquidity is weak in real estate, recoveries concentrate more tightly (conditional on the mean).

In the beta model, we interpret the coefficients on the precision parameter $\phi$. Negative coefficients imply lower precision (and thus higher variance). For example, *Days to Maturity * Energy* (-1.53), *Global PD * Real Estate* (-4.34), and *Liquidity Ratio * Senior Unsecured* (-0.70) are negative, indicating that recoveries in the Energy and Real Estate sectors, as well as for liquid senior unsecured bonds, are inherently more volatile (less precise) than the baseline. Conversely, *S&P500* is positive (+1.04), indicating that stronger market conditions are associated with higher precision (lower dispersion) and more predictable outcomes.

The selection patterns themselves differ in emphasis. The logit-heteroskedastic fit tends to pick sector-dummy interactions (e.g., with *Communications* or *Health Care*), highlighting cross-industry shifts in how fundamentals map into recoveries. The beta fit more often chooses horizon/level-specific interactions (e.g., *Days to Maturity * Energy*, *Prior PD * Real Estate*), emphasizing timing and near-default firm condition.

## 4.5 Conclusion

This paper develops a flexible framework for modelling post-default recovery rates that casts $\ell_0$-sparse variable selection as a discrete sampling problem over predictor permutations and solves it with density/probability-tempered Sequential Monte Carlo under a cross-validated objective. The approach is modular—accommodating alternative conditional distributions (here, beta and logit-heteroskedastic), preprocessing choices, and interaction designs—and remains computationally practical even in settings with high-dimensional, correlated predictor sets. A key design feature is the ability to initialize with a comprehensive candidate set, including rich continuous×categorical interactions, while converging to a parsimonious, interpretable model. We find that optional light prescreening further enhances interpretability by reducing redundancy without altering the central economic conclusions.

Empirically, we evaluate four nested designs that progressively enrich the information set, culminating in a saturated specification that includes firm-specific covariates and interactions (*Basic+Firm+Inter.(All)*). As the information set expands, both models demonstrate consistent improvements in out-of-sample fit.

Our primary contribution is utilizing the inherent flexibility of this framework to consistently evaluate competing model specifications. By subjecting both the beta and logit-heteroskedastic models to an identical selection mechanism, we are able to assess their relative strengths across multiple performance dimensions. While the logit-heteroskedastic model achieves higher log-likelihood scores, closer inspection reveals this is largely driven by its fit at the distribution's boundaries (due to the Jacobian transformation). Conversely, the beta model demonstrates superior predictive accuracy ($R^2$) and is the only specification supported by the Conditional Kolmogorov calibration test. Based on this multi-faceted evaluation that prioritizes characterization of the conditional mean and formal specification validity, we identify the beta model as the more suitable specification for this dataset and

empirical setting. The SMC framework thus serves as a valuable tool for conducting such comprehensive model comparisons.

The economic implications derived from the representative model are consistent and intuitive. *Days to Maturity* is positively associated with recovery rates, whereas *Subordinated* status and wider *Yield Spreads* predict lower recoveries. Interactions reveal significant sectoral heterogeneity: recoveries are especially depressed when *Domestic PD* is high in the *Real Estate* sector, whereas *Utilities* appear less exposed to global credit shocks. Furthermore, the dual-equation specification highlights risk dynamics: we find that sectors such as *Energy* and *Real Estate* exhibit negative coefficients in the precision equation, implying significantly higher volatility (lower predictability) in their recovery outcomes compared to the market baseline.

# Chapter 5

# Conclusion and Future Directions

## 5.1 Summary of Findings and Contributions

This thesis has systematically investigated the capacity of SMC methods to serve as a unified computational backbone for modern econometrics. By re-examining the theoretical underpinnings of the Feynman-Kac formalism, we demonstrated that a single inferential engine can effectively resolve two mathematically distinct classes of problems: high-dimensional combinatorial optimization and non-linear parameter inference. The central conclusion of this work is that while SMC provides a powerful generalized framework, it is not a black box solution; its successful application necessitates a rigorous tailoring of the particle system's interpretation and specific tuning of proposal mechanisms to navigate the unique topological challenges of economic models. We established that density-tempered SMC offers a robust mechanism for solving the NP-hard problem of $l_0$-sparse variable selection in credit risk, while a smooth-likelihood particle filter renders Bayesian inference computationally feasible for high-signal, mixed-frequency LRR models.

## 5.2 Synthesis of Empirical Approaches and Challenges

To evaluate the efficacy of these methodologies, two distinct empirical strategies were employed using U.S. financial data, addressing regimes where standard econometric techniques proved insufficient.

In the first application, the estimation of a nonlinear LRR model characterized by recursive preferences and stochastic volatility was addressed. The framework was augmented to include preference shocks, and a specific modelling choice was made to treat conditional variance as an autoregressive gamma process rather than a standard autoregressive specification. This ensured the positivity of variance without ad-hoc truncation and provided a well-defined stationary gamma distribution. The model was cast as a nonlinear state-space system explicitly accommodating mixed-frequency data, utilizing quarterly observations for consumption and dividend growth alongside monthly observations for market returns and risk-free rates.

The primary estimation obstacle in this setting was the computational instability of standard pseudo-marginal pipelines. It was argued that in the presence of nonlinear dynamics and highly informative observations, standard particle filters require prohibitive numbers of evaluations to yield stable likelihood estimates. To circumvent this, the S-UPF-CRN-SMC strategy was implemented. This approach replaced the requirement for unbiased likelihoods with a stable surrogate, achieved by combining a Gaussian smoothing approximation around the inferred state mean with CRNs. This preserved the geometry of the likelihood surface needed for posterior inference while dramatically reducing Monte Carlo noise. Posterior inference was then performed via a tempered SMC sampler that bridged from the prior to the posterior, using the smoothed surrogate to update weights and employing MCMC rejuvenation moves to maintain particle diversity.

In the second application, post-default bond recovery was treated as a high-dimensional model selection problem. The fundamental challenge was to identify a compact, predictive subset of covariates from a large candidate space containing hundreds of predictors, including firm-specific health measures and interaction terms. This combinatorial environment renders convex relaxation methods prone to shrinkage bias and greedy algorithms susceptible to local optima. The central methodological contribution was the reformulation of l0-sparse selection as a discrete sampling problem over predictor subsets, implemented via a tempered Sequential Monte Carlo algorithm.

The algorithm represented candidate models as permutations of selected predictors and transitioned from an initial distribution to the target posterior using an adaptive tempering schedule. A defining feature of this implementation was that model selection was driven by a cross-validated objective function, where each candidate subset was evaluated on held-out folds. This embedded generalization discipline directly into the discrete target

being explored. To prevent the early exclusion of relevant variables, the method applied Metropolis-Hastings move steps, referred to as support boosting, which incorporated variable frequencies from the current population. This common selection engine facilitated a rigorous comparison between two competing recovery specifications: a Beta regression model and a Logit-transformed heteroskedastic model, allowing differences in performance to be attributed to model structure rather than covariate choice.

## 5.3 Key Empirical Results and Implications

The analysis yields robust evidence regarding both model specification and structural economic dynamics, addressing the core research questions posed in the introduction.

In the domain of asset pricing, the mixed-frequency estimation revealed that high-frequency financial data is essential for identifying latent volatility dynamics. The analysis used U.S. data spanning from 1959 to 2022, and the most significant findings emerged from contrasting the pre-COVID sample with the extended COVID-inclusive sample. The inclusion of the crisis period was associated with a sizable decline in the posterior mean of risk aversion and a distinct shift in volatility expectations; volatility was inferred to be higher in scale but lower in persistence. This characterizes the pandemic as a high-impact but transitory risk shock rather than a permanent regime change. Additionally, a structural shift was observed in the correlation between consumption and dividend growth, which flipped from weakly positive in the pre-COVID period to sharply negative during the crisis. These results imply that mixed-frequency filtering allows the model to respond quickly to structural breaks, interpreting rapid market repricing as a reaction to transitory volatility spikes while keeping long-run growth expectations anchored.

In the domain of credit recovery, a systematic divergence was found between likelihood-based fit and predictive utility. In the saturated design, the Beta regression model achieved higher out-of-sample $R^2$ values, whereas the Logit-heteroskedastic model tended to attain higher log-likelihoods. This divergence was interpreted as evidence that likelihood gains in the Logit specification reflected improved density fit at the boundaries rather than improved accuracy of the conditional mean. Crucially, model adequacy was evaluated using the Conditional Kolmogorov test; the Beta model was the only specification not rejected, whereas the Logit-heteroskedastic model was rejected at conventional levels.

Calibration plots reinforced this finding, showing that the Logit model produced overly extreme predictions. The selected models in the richest specification highlighted days to maturity, subordination, and yield spreads as robust drivers of recovery, with interaction terms capturing sector-dependent sensitivities. This supports the conclusion that Beta regression offers a better-calibrated framework for bounded recovery data, even when alternative models appear statistically superior on likelihood metrics.

## 5.4 Limitations and Future Research

While this thesis establishes a unified SMC framework for econometric analysis, specific limitations inherent in the current analysis point toward necessary refinements and broader research avenues.

First, regarding the LRR application, the inference strategy currently relies on a Gaussian approximation of the filtering density. While computationally efficient, this approximation may degrade in regimes exhibiting extreme skewness or multimodality, potentially introducing bias in the estimation of structural parameters during extreme tail events. Future algorithmic refinements could ameliorate this by adopting Gaussian mixture approximations within the filtering step. This would allow the filter to capture the skewness and multimodality of latent states, a capability essential for highly non-linear dynamics, while maintaining the smoothness required for efficient MCMC proposals. Furthermore, moving beyond the current log-linear solution methods toward fully nonlinear global projection approaches would eliminate linearisation errors, providing a more rigorous test of the model's structural restrictions.

In terms of measurement and forecasting, the current equation for dividend growth could be enhanced to accommodate heavier tails or time-varying idiosyncratic volatility, thereby improving the model's robustness to extreme outliers. Additionally, explicit recursive forecasting exercises, involving training strictly on pre-crisis data and testing on crisis windows, would serve to quantify the real-time early-warning capabilities of the mixed-frequency framework.

Second, regarding the credit recovery application, the results are derived exclusively from U.S. corporate bond data. Given the substantial heterogeneity in bankruptcy codes and creditor rights across jurisdictions, the determinants identified here may not generalize to

European or Asian markets. Extending the analysis to an international panel would provide a valuable test of the transitory shock hypothesis. This would allow for a cross-country comparison of how differing fiscal and monetary responses to global shocks, such as the COVID-19 pandemic, altered risk premia dynamics across disparate sovereign bond markets.

Finally, regarding the combinatorial optimization methodology used for variable selection, the current analysis focuses on the internal consistency of the SMC approach. Future research should conduct formal comparison tests against traditional operations research methods, such as linear integer programming. Benchmarking the SMC sampler against exact solvers would help quantify the trade-offs between computational speed and solution optimality in high-dimensional feature spaces.

In conclusion, this thesis confirms that SMC methods are not merely filtering algorithms, but a versatile computational framework capable of solving complex optimization and inference challenges. However, their power lies in their adaptability; successful implementation requires a rigorous tuning of the algorithm to the specific structure of the economic question at hand.

# Bibliography


Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., and Stuart, A. M. (2017). "Importance sampling: Intrinsic dimension and computational cost". *Statistical Science* 32.3, pp. 405–431.

Albuquerque, R., Eichenbaum, M., Luo, V., and Rebelo, S. (2016). "Valuation risk and asset pricing". *Journal of Finance* 71, pp. 2861–2904.

An, S. and Schorfheide, F. (2007). "Bayesian analysis of DSGE models". *Econometric Reviews* 26.2–4, pp. 113–172.

Andrews, D. W. K. (1997). "A conditional Kolmogorov test". *Econometrica*, pp. 1097–1128.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). "Particle Markov chain Monte Carlo methods". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3, pp. 269–342.

Andrieu, C. and Roberts, G. O. (2009). "The pseudo-marginal approach for efficient Monte Carlo computations". *The Annals of Statistics* 37.2, pp. 697–725.

Bansal, R., Gallant, A. R., and Tauchen, G. (2007). "Rational pessimism, rational exuberance, and asset pricing models". *The Review of Economic Studies* 74.4, pp. 1005–1033.

Bansal, R., Kiku, D., and Yaron, A. (2016). "Risks for the long run: Estimation with time aggregation". *Journal of Monetary Economics* 82, pp. 52–69.

Bansal, R. and Yaron, A. (2004). "Risks for the long run: A potential resolution of asset pricing puzzles". *Journal of Finance* 59.4, pp. 1481–1509.

Beeler, J. and Campbell, J. Y. (2009). *The long-run risks model and aggregate asset prices: An empirical assessment*. Tech. rep. 14788. National Bureau of Economic Research.

Bellotti, T. and Crook, J. (2012). "Loss given default models incorporating macroeconomic variables for credit cards". *International Journal of Forecasting* 28.1, pp. 171–182.

Calabrese, R. (2014). "Predicting bank loan recovery rates with a mixed continuous–discrete model". *Applied Stochastic Models in Business and Industry* 30.2, pp. 99–114.

Calvet, Laurent E. and Czellar, Veronika (2015). "Through the Looking Glass: Indirect Inference via Simple Equilibria". *Journal of Econometrics* 185.2, pp. 343–358.

Campbell, J. Y. and Shiller, R. J. (1988). "The dividend-price ratio and expectations of future dividends and discount factors". *The Review of Financial Studies* 1.3, pp. 195–228.

Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in Hidden Markov Models*. New York: Springer.

Chopin, N. (2002). "A sequential particle filter method for static models". *Biometrika* 89, pp. 539–551.

Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). "SMC$^2$: An efficient algorithm for sequential analysis of state space models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.3, pp. 397–426.

Chopin, N. and Papaspiliopoulos, O. (2020). *An Introduction to Sequential Monte Carlo*. Springer.

Creal, D. (2012). "A survey of sequential Monte Carlo methods for economics and finance". *Econometric Reviews* 31.3, pp. 245–296.

Dai, C., Heng, J., Jacob, P. E., and Whiteley, N. (2022). "An invitation to sequential Monte Carlo samplers". *Journal of the American Statistical Association* 117.539, pp. 1587–1600.

Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Vol. 88. New York: Springer.

Del Moral, P., Doucet, A., and Jasra, A. (2006). "Sequential Monte Carlo samplers". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.3, pp. 411–436.

Deligiannidis, G., Doucet, A., Pitt, M. K., and Kohn, R. (2018). "The correlated pseudo-marginal method". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.5, pp. 839–870.

Diaconis, P. (2009). "The Markov chain Monte Carlo revolution". *Bulletin of the American Mathematical Society* 46.2, pp. 179–205.

Doucet, A., Freitas, N. de, and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer.

Doucet, A., Godsill, S., and Andrieu, C. (2000). "On sequential Monte Carlo sampling methods for Bayesian filtering". *Statistics and Computing* 10.3, pp. 197–208.

Doucet, A. and Johansen, A. M. (2009). "A tutorial on particle filtering and smoothing: Fifteen years later". *Handbook of Nonlinear Filtering* 12, pp. 656–704.

Doucet, A., Pitt, M. K., Deligiannidis, G., and Kohn, R. (2015). "Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator". *Biometrika* 102.2, pp. 295–313.

Duan, J. C. (2019). *Variable Selection with Big Data based on Zero Norm and via Sequential Monte Carlo*. Available at SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3377038.

Duan, J. C. and Fulop, A. (2015). "Density-tempered marginalized sequential Monte Carlo samplers". *Journal of Business & Economic Statistics* 33.2, pp. 192–202.

Duan, J. C. and Hwang, R. C. (2014). *Predicting recovery rate at the time of corporate default*. Working paper.

Duan, J. C., Li, S., and Xu, Y. (2023). "Sequential Monte Carlo optimization and statistical inference". *Wiley Interdisciplinary Reviews: Computational Statistics* 15.3, e1598.

Duan, J. C., Sun, J., and Wang, T. (2012). "Multiperiod corporate default prediction—A forward intensity approach". *Journal of Econometrics* 170.1, pp. 191–209.

Duan, J. C. and Zhang, C. (2016). *Non-Gaussian bridge sampling with an application*. Available at SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2675877.

Duffie, D., Saita, L., and Wang, K. (2007). "Multi-period corporate default prediction with stochastic covariates". *Journal of Financial Economics* 83.3, pp. 635–665.

Epstein, L. G. and Zin, S. E. (1989). "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework". *Econometrica* 57, pp. 937–969.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2005). "Estimating dynamic equilibrium economies: linear versus nonlinear likelihood". *Journal of Applied Econometrics* 20, pp. 891–910.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). "Estimating macroeconomic models: a likelihood approach". *The Review of Economic Studies* 74.4, pp. 1059–1087.

Fulop, A., Heng, J., Li, J., and Liu, H. (2022a). "Bayesian estimation of long-run risk models using sequential Monte Carlo". *Journal of Econometrics* 228.1, pp. 62–84.

Fulop, A. and Li, J. (2013). "Efficient learning via simulation: A marginalized resample-move approach". *Journal of Econometrics* 176.2, pp. 146–161.

Fulop, A. and Li, J. (2019). "Bayesian estimation of dynamic asset pricing models with informative observations". *Journal of Econometrics* 209.1, pp. 114–138.

Fulop, A., Li, J., Liu, H., and Yan, C. (2024). "Estimating and Testing Long-Run Risk Models: International Evidence". *Management Science* 70.7, pp. 4564–4603.

Fulop, Andras, Heng, Jeremy, and Li, Junye (2022b). *Efficient Likelihood-based Estimation via Annealing for Dynamic Structural Macrofinance Models*. DOI: arXiv.2201.01094.

George, E. I. and McCulloch, R. E. (1993). "Variable selection via Gibbs sampling". *Journal of the American Statistical Association* 88.423, pp. 881–889.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". *IEE Proceedings F–Radar and Signal Processing*. Vol. 140. 2. IET, pp. 107–113.

Gourieroux, C. and Jasiak, J. (2006). "Autoregressive gamma processes". *Journal of Forecasting* 25.2, pp. 129–152.

Grisetti, G., Stachniss, C., and Burgard, W. (2007). "Improved techniques for grid mapping with Rao-Blackwellized particle filters". *IEEE Transactions on Robotics* 23.1, pp. 34–46.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications". *Biometrika* 57.1, pp. 97–109.

Heng, J., Bishop, A. N., Deligiannidis, G., and Doucet, A. (2020). "Controlled sequential Monte Carlo". *The Annals of Statistics* 48.5, pp. 2904–2929.

Herbst, E. P. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models*. Princeton, NJ: Princeton University Press.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., and Chopin, N. (2015). "On particle methods for parameter estimation in state-space models". *Statistical Science* 30.3, pp. 328–351.

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic volatility: Likelihood inference and comparison with ARCH models". *The Review of Economic Studies* 65.3, pp. 361–393.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). "Equation of state calculations by fast computing machines". *The Journal of Chemical Physics* 21.6, pp. 1087–1092.

Mitchell, T. J. and Beauchamp, J. J. (1988). "Bayesian variable selection in linear regression". *Journal of the American Statistical Association* 83.404, pp. 1023–1032.

Murphy, K. and Russell, S. (2001). "Rao-Blackwellised particle filtering for dynamic Bayesian networks". *Sequential Monte Carlo Methods in Practice*. Ed. by Doucet, A., Freitas, J. F. G. de, and Gordon, N. J. New York, NY: Springer, pp. 499–515.

Natarajan, B. K. (1995). "Sparse approximate solutions to linear systems". *SIAM Journal on Computing* 24.2, pp. 227–234.

Neal, R. M. (2001). "Annealed importance sampling". *Statistics and Computing* 11, pp. 125–139.

Neal, R. M. (2011). "MCMC using Hamiltonian dynamics". *Handbook of Markov Chain Monte Carlo*. Ed. by Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. Chapman & Hall/CRC, pp. 113–162.

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rad, K. R., Vidne, M., Vogelstein, J., and Wu, W. (2010). "A new look at state-space models for neural data". *Journal of Computational Neuroscience* 29, pp. 107–126.

Pitt, M. K., Santos Silva, R. dos, Giordani, P., and Kohn, R. (2012). "On some properties of Markov chain Monte Carlo simulation methods based on the particle filter". *Journal of Econometrics* 171.2, pp. 134–151.

Pitt, M. K. and Shephard, N. (1999). "Filtering via simulation: Auxiliary particle filters". *Journal of the American Statistical Association* 94.446, pp. 590–599.

Polson, N. G. and Xu, Y. (2018). "Deep learning: a Bayesian perspective". *Bayesian Analysis* 13.4, pp. 1275–1304.

Roberts, G. O. and Tweedie, R. L. (1996). "Exponential convergence of Langevin distributions and their discrete approximations". *Bernoulli* 2.4, pp. 341–363.

Schäfer, C. and Chopin, N. (2013). "Sequential Monte Carlo on large binary sampling spaces". *Statistics and Computing* 23.2, pp. 163–184.

Schorfheide, F., Song, D., and Yaron, A. (2018). "Identifying long-run risks: A Bayesian mixed-frequency approach". *Econometrica* 86, pp. 617–654.

Sherlock, C., Thiery, A. H., Roberts, G. O., and Rosenthal, J. S. (2015). "On the efficiency of pseudo-marginal random walk Metropolis algorithms". *The Annals of Statistics* 43.1, pp. 238–275.

Soussen, C., Idier, J., Duan, J., and Brie, D. (2015). "Homotopy Based Algorithms for $\ell_0$-Regularized Least-Squares". *IEEE Transactions on Signal Processing* 63.13, pp. 3301–3316.

Van der Merwe, R., Doucet, A., De Freitas, N., and Wan, E. A. (2000). "The Unscented particle Filter". *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. IEEE, pp. 153–158.

Weil, P. (1989). "The equity premium puzzle and the risk-free return puzzle". *Journal of Monetary Economics* 24, pp. 401–421.

Welch, G. and Bishop, G. (1995). *An introduction to the Kalman filter*. Technical Report. University of North Carolina at Chapel Hill.

Wilkinson, D. J. (2012). *Stochastic Modelling for Systems Biology*. 2nd. Boca Raton, FL: CRC Press.

Zhou, Y., Johansen, A. M., and Aston, J. A. D. (2016). "Toward automatic model comparison: An adaptive sequential Monte Carlo approach". *Journal of Computational and Graphical Statistics* 25.3, pp. 701–726.

Zviadadze, I. (2017). "Term structure of consumption risk premia in the cross-section of currency returns". *Journal of Finance* 72, pp. 1529–1578.

# Appendix A

# Appendix to Chapters 2

**Algorithm 1** Metropolis-Hastings Algorithm

**Input:** Initial state $x_0$, target distribution $\pi(x)$, proposal density $q(x'|x)$, iterations $T$.
**for** $t = 1$ to $T$ **do**
    **1. Propose:** Generate candidate $x^* \sim q(\cdot|x_{t-1})$.
    **2. Ratio:** Calculate acceptance probability $\alpha$:

$$\alpha = \min\left\{1, \frac{\pi(x^*)\, q(x_{t-1}|x^*)}{\pi(x_{t-1})\, q(x^*|x_{t-1})}\right\}$$

    **3. Acceptance:** Sample $u \sim \mathcal{U}[0, 1]$.
    **if** $u \leq \alpha$ **then**
        $x_t \leftarrow x^*$ ▷ Accept move
    **else**
        $x_t \leftarrow x_{t-1}$ ▷ Reject move
**Output:** Markov chain samples $\{x_1, \ldots, x_T\}$.

**Algorithm 2** MALA Kernel ($K_t$ for one particle $x$)

**Input:** Current state $x$, target $\pi \propto \gamma$, step size $\epsilon$, preconditioner $G$.
**1. Propose:** Generate candidate $x'$:

$$x' \sim \mathcal{N}\left(x + \frac{\epsilon^2}{2} G^{-1} \nabla \log \gamma(x),\, \epsilon^2 G^{-1}\right)$$

**2. Ratio:** Calculate transition probabilities $q(x'|x)$ and $q(x|x')$ using proposal density.
**3. Acceptance:** Calculate probability $\alpha$:

$$\alpha = \min\left\{1, \frac{\gamma(x')\, q(x|x')}{\gamma(x)\, q(x'|x)}\right\}$$

**4. Update:** Sample $u \sim \mathcal{U}[0, 1]$.
**if** $u \leq \alpha$ **then**
    **return** $x'$
**else**
    **return** $x$

**Algorithm 3** HMC Kernel ($K_t$ for one particle $x$)

**Input:** Current state $x$, potential $U(x) = -\log \gamma(x)$, step size $\epsilon$, steps $L$, mass matrix $M$.
**1. Initialize:** Sample momentum $p \sim \mathcal{N}(0, M)$. Set $(x^{(0)}, p^{(0)}) = (x, p)$.
**2. Leapfrog Integrator:**
  $p^{(1/2)} = p^{(0)} - \frac{\epsilon}{2} \nabla U(x^{(0)})$ ▷ Half-step momentum
**for** $l = 0$ to $L - 1$ **do**
    $x^{(l+1)} = x^{(l)} + \epsilon M^{-1} p^{(l+1/2)}$ ▷ Full-step position
    **if** $l < L - 1$ **then**
      $p^{(l+3/2)} = p^{(l+1/2)} - \epsilon \nabla U(x^{(l+1)})$ ▷ Full-step momentum
  $p^{(L)} = p^{(L-1/2)} - \frac{\epsilon}{2} \nabla U(x^{(L)})$ ▷ Final half-step
**3. Metropolis Correction:** Set proposal $(x', p') = (x^{(L)}, -p^{(L)})$.

$$\alpha = \min\left\{1, \exp\left(-H(x', p') + H(x^{(0)}, p^{(0)})\right)\right\}$$

**4. Update:** If $u \leq \alpha$, return $x'$; else return $x^{(0)}$.

**Algorithm 4** Bootstrap Particle Filter (BPF)

**Input:** Particles $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}$ and observation $y_t$.

**1. Resample:** If $\text{ESS}_{t-1} < N/2$, resample ancestors $\{\tilde{x}_{t-1}^{(i)}\}$ from $\{x_{t-1}^{(j)}\}$ with probabilities $w_{t-1}$. Reset weights $w_{t-1}^{(i)} = 1/N$. Otherwise, set $\tilde{x}_{t-1}^{(i)} = x_{t-1}^{(i)}$.

**2. Propagate:** For $i = 1$ to $N$, sample from transition:

$$x_t^{(i)} \sim f_\theta(\cdot|\tilde{x}_{t-1}^{(i)})$$

**3. Weight:** Compute unnormalized weights:

$$\omega_t^{(i)} = w_{t-1}^{(i)} \cdot g_\theta(y_t|x_t^{(i)})$$

**4. Normalize:** $w_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^{N} \omega_t^{(j)}$.

**Algorithm 5** Auxiliary Particle Filter (APF)

**Input:** Particles $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}$ and observation $y_t$.

**1. Look-ahead Scoring:** Compute predictive score $\mu_t^{(i)} \approx p(y_t|x_{t-1}^{(i)})$.

**2. Auxiliary Weights:** Compute pre-weights $\tilde{\omega}_t^{(i)} \propto w_{t-1}^{(i)} \cdot \mu_t^{(i)}$.

**3. Ancestor Selection:** Sample indices $a_i \sim \text{Categorical}(\tilde{\omega}_t)$. Set $\tilde{x}_{t-1}^{(i)} = x_{t-1}^{(a_i)}$.

**4. Propagate:** Sample $x_t^{(i)} \sim q_t(\cdot|\tilde{x}_{t-1}^{(i)}, y_t)$.

**5. Weight Correction:**

$$\omega_t^{(i)} = \frac{g_\theta(y_t|x_t^{(i)})\, f_\theta(x_t^{(i)}|\tilde{x}_{t-1}^{(i)})}{q_t(x_t^{(i)}|\tilde{x}_{t-1}^{(i)}, y_t)\, \mu_t^{(a_i)}}$$

**6. Normalize:** $w_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^{N} \omega_t^{(j)}$.

**Algorithm 6** Unscented Particle Filter (UPF) - Core Step

**Input:** Particles $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}$ and observation $y_t$.
**1. Resample:** If $\text{ESS}_{t-1} < N/2$, resample to get $\{\tilde{x}_{t-1}^{(i)}\}$.
**for** $i = 1$ to $N$ **do**
    **2. UKF Step:** Initialize UKF with mean $\tilde{x}_{t-1}^{(i)}$ and covariance $P_{t-1}$.
        a. Generate Sigma Points $\mathcal{X}_{t-1}^{(i)}$.
        b. Predict State via $f_\theta$ to get $\hat{x}_{t|t-1}^{(i)}, P_{t|t-1}^{(i)}$.
        c. Predict Obs via $g_\theta$ to get $\hat{y}_t^{(i)}, S_t^{(i)}$.
        d. Update mean/cov to $(\mu_{t|t}^{(i)}, \Sigma_{t|t}^{(i)})$ using Kalman gain.
    **3. Propagate:** Sample using optimal Gaussian approximation:

$$x_t^{(i)} \sim \mathcal{N}(\mu_{t|t}^{(i)}, \Sigma_{t|t}^{(i)})$$

    **4. Weight:**

$$\omega_t^{(i)} = \frac{g_\theta(y_t|x_t^{(i)})\, f_\theta(x_t^{(i)}|\tilde{x}_{t-1}^{(i)})}{\mathcal{N}(x_t^{(i)}; \mu_{t|t}^{(i)}, \Sigma_{t|t}^{(i)})}$$

**5. Normalize:** $w_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^{N} \omega_t^{(j)}$.

**Algorithm 7** Rao-Blackwellised Particle Filter (RBPF) - Core Step

**Input:** Particles $\{x_{t-1,nl}^{(i)}, \mu_{t-1,l}^{(i)}, P_{t-1,l}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}$.
**if** Resampling criterion met **then**
    Resample particles (incl. their KF states $\mu_{t-1,l}, P_{t-1,l}$); Reset $w_{t-1}^{(i)} = 1/N$.
**for** $i = 1$ to $N$ **do**
    **Propagate (Nonlinear):** Sample $x_{t,nl}^{(i)} \sim q_t(\cdot \mid x_{t-1,nl}^{(i)}, y_t)$. ▷ Proposal for nl part
    **Update (Linear KF):** Run one step of the KF for the linear state $x_{t,l}$ *conditional on* $x_{t,nl}^{(i)}$ and $x_{t-1,nl}^{(i)}$. Obtain updated statistics $\mu_{t,l}^{(i)}, P_{t,l}^{(i)}$ and the conditional likelihood $p_\theta(y_t \mid y_{1:t-1}, x_{1:t,nl}^{(i)})$.
    **Weight:** Compute unnormalized weight $\omega_t^{(i)} = w_{t-1}^{(i)} \cdot \frac{p_\theta(y_t|y_{1:t-1}, x_{1:t,nl}^{(i)})\, p_\theta(x_{t,nl}^{(i)}|x_{t-1,nl}^{(i)})}{q_t(x_{t,nl}^{(i)}|x_{t-1,nl}^{(i)}, y_t)}$.
**Normalize:** $w_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^{N} \omega_t^{(j)}$.
**Output:** Particles $\{x_{t,nl}^{(i)}, \mu_{t,l}^{(i)}, P_{t,l}^{(i)}, w_t^{(i)}\}_{i=1}^{N}$.

**Algorithm 8** Particle Metropolis-Hastings (PMMH)

**Initialization:** Sample $\theta^{(0)} \sim p(\theta)$. Run PF to get $\hat{p}(y_{1:T}|\theta^{(0)})$.
**Iteration:** For $i = 1$ to $M$:
**1. Propose:** Sample candidate $\theta^* \sim q(\cdot|\theta^{(i-1)})$.
**2. Particle Filter:** Run PF with $N_x$ particles using $\theta^*$ to get $\hat{p}(y_{1:T}|\theta^*)$.
**3. Acceptance Probability:**

$$\alpha = \min\left\{1, \frac{\hat{p}(y_{1:T}|\theta^*)\, p(\theta^*)\, q(\theta^{(i-1)}|\theta^*)}{\hat{p}(y_{1:T}|\theta^{(i-1)})\, p(\theta^{(i-1)})\, q(\theta^*|\theta^{(i-1)})}\right\}$$

**4. Accept/Reject:** Sample $u \sim \mathcal{U}[0,1]$.
**if** $u \leq \alpha$ **then**
    Set $\theta^{(i)} = \theta^*$.
**else**
    Set $\theta^{(i)} = \theta^{(i-1)}$.

**Algorithm 9** SMC$^2$ Algorithm (Data Tempering)

**Initialization:** Sample $\{\theta_0^{(j)}\}_{j=1}^{N_\theta} \sim p(\theta)$. Initialize inner particle systems $S_0^{(j)}$ with $1/N_x$ weights.
**Iteration ($t = 1$ to $T$):**
**for** $j = 1$ to $N_\theta$ **do**
    **1. Propagate Inner PF:** Update $S_{t-1}^{(j)} \to S_t^{(j)}$ with $y_t$. Get incremental likelihood estimate $\hat{p}(y_t|y_{1:t-1}, \theta_{t-1}^{(j)})$.
    **2. Update Outer Weights:** $\omega_t^{(j)} = W_{t-1}^{(j)} \cdot \hat{p}(y_t|y_{1:t-1}, \theta_{t-1}^{(j)})$.
**3. Normalize:** $W_t^{(j)} = \omega_t^{(j)} / \sum_{k=1}^{N_\theta} \omega_t^{(k)}$.
**4. Resample (Outer):** If $\mathrm{ESS}_t < \kappa N_\theta$, resample $\{\theta_{t-1}^{(j)}, S_t^{(j)}\}$.
**5. Rejuvenate (Move):** Apply PMMH kernel to each $\theta_t^{(j)}$ targeting $p(\theta|y_{1:t})$.

**Algorithm 10** SMC$^2$ Algorithm (Density Tempering)

**Initialization ($k = 0$):**
Sample parameter particles $\{\theta_0^{(j)}\}_{j=1}^{N_\theta} \sim p(\theta)$. Set $\lambda_0 = 0$.
For each $j$, run a full Particle Filter (with $N_x$ particles) on data $y_{1:T}$ to estimate the likelihood $\hat{L}^{(j)} = \hat{p}(y_{1:T}|\theta_0^{(j)})$.
Set weights $W_0^{(j)} = 1/N_\theta$.
**Iteration:**
**while** $\lambda_{k-1} < 1$ **do**
    $k \leftarrow k + 1$.
    **1. Select Temperature:** Choose $\lambda_k \in (\lambda_{k-1}, 1]$ adaptively (e.g., to maintain ESS $\approx 0.5N_\theta$).
    **2. Reweight:** Compute unnormalized weights based on the temperature increment:

$$w_k^{(j)} = W_{k-1}^{(j)} \cdot \left(\hat{L}^{(j)}\right)^{\lambda_k - \lambda_{k-1}}$$

    **3. Normalize:** $W_k^{(j)} = w_k^{(j)} / \sum_{m=1}^{N_\theta} w_k^{(m)}$.
    **4. Resample:** If $\text{ESS}_k < \kappa N_\theta$, resample particles $\{\theta_{k-1}^{(j)}, \hat{L}^{(j)}\}$ according to $W_k$. Reset weights to $1/N_\theta$.
    **5. Rejuvenate (Move):**
    **for** $j = 1$ to $N_\theta$ **do**
        Apply a PMMH kernel to $\theta_k^{(j)}$ targeting the tempered posterior:

$$\pi_{\lambda_k}(\theta) \propto p(\theta) \cdot [\hat{p}(y_{1:T}|\theta)]^{\lambda_k}$$

        ▷ If a new $\theta^*$ is accepted, update its stored likelihood $\hat{L}^{(j)}$.

**Algorithm 11** Tempered SMC for Optimization/Selection

**Initialization ($t = 0$):** Set $\lambda_0 = 0$. Sample $x_0^{(i)} \sim \pi_0(x)$. Set $W_0^{(i)} = 1/N$.
**Iteration:**
**while** $\lambda_{t-1} < 1$ **do**
    $t \leftarrow t + 1$.
    **1. Select Temperature:** Choose $\lambda_t \in (\lambda_{t-1}, 1]$ adaptively (e.g., target ESS).
    **2. Reweight:** For each particle $i = 1 \dots N$:

$$w_t^{(i)} = W_{t-1}^{(i)} \cdot \left( \frac{\gamma_T(x_{t-1}^{(i)})}{\pi_0(x_{t-1}^{(i)})} \right)^{\lambda_t - \lambda_{t-1}}$$

    **3. Normalize:** $W_t^{(i)} = w_t^{(i)} / \sum_{j=1}^N w_t^{(j)}$.
    **4. Resample:** If $\mathrm{ESS}_t < \kappa N$, resample.
    **5. Rejuvenate (Move):** Apply MCMC kernel $K_{\lambda_t}$ to $x_t^{(i)}$.

# Appendix B

# Appendix to Chapter 3

## B.1 Derivation of the Autoregressive-gamma Transition

The time evolution of $\sigma_t^2$ is modelled as an autoregression gamma (ARG) process:

$$\sigma_t^2 \sim \mathcal{G}(\phi_s + \epsilon_t, c), \quad \epsilon_t \sim \mathcal{P}(\nu\sigma_{t-1}^2/c),$$

where $\phi_s \in (1, \infty)$, $c \in (0, \infty)$ and $\nu \in (0, 1)$.

We will write

$$\mathcal{G}(s; a, b) = \frac{1}{\Gamma(a)b^a} s^{a-1} \exp(-s/b), \quad s \in (0, \infty),$$

to denote a Gamma distribution with shape $a \in (0, \infty)$ and scale $b \in (0, \infty)$, and

$$\mathcal{P}(\epsilon; \lambda) = \frac{\lambda^\epsilon \exp(-\lambda)}{\epsilon!}, \quad \epsilon \in \mathbb{N}_0,$$

to denote a Poisson distribution with rate $\lambda \in (0, \infty)$, respectively. By integrating out $\epsilon_t$,

the marginal Markov transition density of B.1 is

$$\begin{aligned} f_\theta(\sigma_t^2|\sigma_{t-1}^2) &= \sum_{\epsilon_t=0}^{\infty} \mathcal{P}(\epsilon_t; \nu\sigma_{t-1}^2/c)\mathcal{G}(\sigma_t^2; \phi_s + \epsilon_t, c) \\ &= \sum_{\epsilon_t=0}^{\infty} \frac{(\nu\sigma_{t-1}^2/c)^{\epsilon_t} \exp(-\nu\sigma_{t-1}^2/c)}{\epsilon_t!} \frac{(\sigma_t^2)^{\phi_s+\epsilon_t-1} \exp(-\sigma_t^2/c)}{\Gamma(\phi_s+\epsilon_t)c^{\phi_s+\epsilon_t}} \\ &= \exp\left(-\frac{(\sigma_t^2 + \nu\sigma_{t-1}^2)}{c}\right)(\sigma_t^2)^{\phi_s-1}\frac{1}{c^{\phi_s}} \sum_{\epsilon_t=0}^{\infty} \frac{(\nu\sigma_{t-1}^2/c)^{\epsilon_t}}{\epsilon_t!}\frac{(\sigma_t^2)^{\epsilon_t}}{\Gamma(\phi_s+\epsilon_t)c^{\epsilon_t}}. \end{aligned}$$

By writing the sum as

$$\begin{aligned} &\sum_{\epsilon_t=0}^{\infty} \frac{(\nu\sigma_{t-1}^2/c)^{\epsilon_t}}{\epsilon_t!}\frac{(\sigma_t^2)^{\epsilon_t}}{\Gamma(\phi_s+\epsilon_t)c^{\epsilon_t}} \\ &= \nu^{-(\phi_s-1)/2}\left(\frac{\sigma_{t-1}\sigma_t}{c}\right)^{-(\phi_s-1)} \nu^{(\phi_s-1)/2}\left(\frac{\sigma_{t-1}^2\sigma_t^2}{c^2}\right)^{\phi_s-1} \sum_{\epsilon_t=0}^{\infty} \frac{(\nu\sigma_{t-1}^2\sigma_t^2/c^2)^{\epsilon_t}}{\epsilon_t!\Gamma((\phi_s-1)+\epsilon_t+1)} \\ &= \nu^{-(\phi_s-1)/2}\left(\frac{\sigma_{t-1}\sigma_t}{c}\right)^{-(\phi_s-1)} I_{\phi_s-1}\left(\frac{2\nu^{1/2}\sigma_{t-1}\sigma_t}{c}\right), \end{aligned}$$

where

$$I_\zeta(x) = (x/2)^\zeta \sum_{i=0}^{\infty} \frac{(x^2/4)^i}{i!\Gamma(\zeta+i+1)}$$

denotes a modified Bessel function of the first kind, we have

$$\begin{aligned} f_\theta(\sigma_t^2|\sigma_{t-1}^2) &= \exp\left(-\frac{(\sigma_t^2 + \nu\sigma_{t-1}^2)}{c}\right)(\sigma_t^2)^{\phi_s-1}\frac{1}{c^{\phi_s}}\nu^{-(\phi_s-1)/2}\left(\frac{\sigma_{t-1}\sigma_t}{c}\right)^{-(\phi_s-1)} I_{\phi_s-1}\left(\frac{2\nu^{1/2}\sigma_{t-1}\sigma_t}{c}\right) \\ &= \left(\frac{\sigma_t^2}{\nu\sigma_{t-1}^2}\right)^{(\phi_s-1)/2}\frac{1}{c}\exp\left(-\frac{(\sigma_t^2 + \nu\sigma_{t-1}^2)}{c}\right) I_{\phi_s-1}\left(\frac{2\sqrt{\nu\sigma_{t-1}^2\sigma_t^2}}{c}\right), \end{aligned}$$

which can be seen as a noncentral Gamma distribution. We will write

$$\mathcal{NG}(s; \lambda, a, b) = \sum_{\epsilon=0}^{\infty} \mathcal{P}(\epsilon; \lambda/b)\mathcal{G}(s; a+\epsilon, b), \quad s \in (0, \infty),$$

to denote a noncentral Gamma density with parameters $\lambda, a, b \in (0, \infty)$. It can be shown

that the conditional expectation and variance under $f_\theta(\sigma_{t-1}^2, \sigma_t^2)$ are

$$\mathbb{E}\left[\sigma_t^2|\sigma_{t-1}^2\right] = \phi_s c + \nu\sigma_{t-1}^2, \quad \text{Var}\left[\sigma_t^2|\sigma_{t-1}^2\right] = \phi_s c^2 + 2c\nu\sigma_{t-1}^2,$$

and the stationary distribution of (B.1) is $\mathcal{G}(\phi_s, c/(1-\nu))$. From the latter, the long-run mean is $\bar{\sigma}^2 = \phi_s c/(1-\nu)$. We will initialize the latent process at $s_0 = (x_0, \sigma_0^2) = (0, \bar{\sigma}^2)$. The Markov transition kernel of the latent process has the form

$$f_\theta(\sigma_t^2|\sigma_{t-1}) = \mathcal{NG}(\sigma_t^2; \nu\sigma_{t-1}^2, \phi_s, c)dx_t d\sigma_t^2.$$

### B.1.1 Proposal distributions

As it is not convenient to work with

$$f_\theta(\sigma_t^2|\sigma_{t-1}^2) = \mathcal{NG}(s; \nu\sigma_{t-1}^2, \phi_s, c),$$

we will consider a proposal transition that is given by approximating it with

$$q(\sigma_t^2|\sigma_{t-1}^2, \theta) = \mathcal{LN}(\sigma_t^2; m_q(\sigma_{t-1}^2), V_q(\sigma_{t-1}^2)),$$

where $\mathcal{LN}(s; m, V)$ denotes a log-Normal distribution with parameters $m$ and $V$.

We set

$$\begin{aligned} m_q(\sigma_{t-1}^2) &= \log m_f(\sigma_{t-1}^2) - \frac{V_q(\sigma_{t-1}^2)}{2}, \\ V_q(\sigma_{t-1}^2) &= \log\left(\exp\left(\log V_f(\sigma_{t-1}^2) - 2\log m_f(\sigma_{t-1}^2)\right) + 1\right), \end{aligned}$$

to match the ARG conditional mean $m_f(\sigma_{t-1}^2) = \phi_s c + \nu\sigma_{t-1}^2$ and conditional variance $V_f(\sigma_{t-1}^2) = \phi_s c^2 + 2c\nu\sigma_{t-1}^2$.

By keeping the same latent transition for log-consumption growth, the Markov transition kernel of our proposal process has the form

$$q(\sigma_t^2|\sigma_{t-1}^2, \theta) = \mathcal{LN}\left(\sigma_t^2; m_q(\sigma_{t-1}^2), V_q(\sigma_{t-1}^2)\right).$$

## B.2 Observation Density of the Balanced-Frequency Model

In this section, we assume that all observations arrive at quarterly frequency. Following the previous sections, we have the following quarterly consumption and dividend observations

$$\begin{aligned}
\Delta d^q_{t+3} &= \frac{1}{3}(\mu_d + \Phi x_{t+2} + \phi_{dc}\sigma_{t+2}\eta_{c,t+3} + \phi_d\sigma_{t+2}\eta_{d,t+3}) \\
&\quad + \frac{2}{3}(\mu_d + \Phi x_{t+1} + \phi_{dc}\sigma_{t+1}\eta_{c,t+2} + \phi_d\sigma_{t+1}\eta_{d,t+2}) \\
&\quad + (\mu_d + \Phi x_t + \phi_{dc}\sigma_t\eta_{c,t+1} + \phi_d\sigma_t\eta_{d,t+1}) + \frac{2}{3}\Delta d_t + \frac{1}{3}\Delta d_{t-1} \\
\Delta c^q_{t+3} &= \frac{1}{3}\Delta c_{t+3} + \frac{2}{3}\Delta c_{t+2} + \Delta c_{t+1} + \frac{2}{3}\Delta c_t + \frac{1}{3}\Delta c_{t-1} \\
&= \frac{1}{3}(\mu + x_{t+2} + \sigma_{t+2}\eta_{c,t+3}) + \frac{2}{3}(\mu + x_{t+1} + \sigma_{t+1}\eta_{c,t+2}) \\
&\quad + (\mu + x_t + \sigma_t\eta_{c,t+1}) + \frac{2}{3}\Delta c_t + \frac{1}{3}\Delta c_{t-1}.
\end{aligned}$$

Furthermore, we assume the following for the quarterly market return

$$\begin{aligned}
r^q_{m,t+3} &= r_{m,t+3} + r_{m,t+2} + r_{m,t+1} \\
&= \xi_\theta(z_{t+3}, z_{t+2}) + \xi_\theta(z_{t+2}, z_{t+1}) + \xi_\theta(z_{t+1}, z_t) + \phi_m\eta_{m,t+3} + \phi_m\eta_{m,t+2} \\
&\quad + \phi_m\eta_{m,t+1} + \Delta d_{t+2} + \Delta d_{t+1} + \Delta d_t \\
&= \xi_\theta(z_{t+3}, z_{t+2}) + \xi_\theta(z_{t+2}, z_{t+1}) + \xi_\theta(z_{t+1}, z_t) + \phi_m\eta_{m,t+3} + \phi_m\eta_{m,t+2} + \phi_m\eta_{m,t+1} \\
&\quad + (\mu_d + \Phi x_{t+2} + \phi_{dc}\sigma_{t+2}\eta_{c,t+3} + \phi_d\sigma_{t+2}\eta_{d,t+3}) \\
&\quad + (\mu_d + \Phi x_{t+1} + \phi_{dc}\sigma_{t+1}\eta_{c,t+2} + \phi_d\sigma_{t+1}\eta_{d,t+2}) \\
&\quad + (\mu_d + \Phi x_t + \phi_{dc}\sigma_t\eta_{c,t+1} + \phi_d\sigma_t\eta_{d,t+1}) \\
&= \xi_\theta(z_{t+3}, z_{t+2}) + \xi_\theta(z_{t+2}, z_{t+1}) + \xi_\theta(z_{t+1}, z_t) + 3\mu_d + \Phi(x_{t+2} + x_{t+1} + x_t) \\
&\quad + \phi_m\eta_{m,t+3} + \phi_m\eta_{m,t+2} + \phi_m\eta_{m,t+1} + \phi_{dc}\sigma_{t+2}\eta_{c,t+3} + \phi_{dc}\sigma_{t+1}\eta_{c,t+2} + \phi_{dc}\sigma_t\eta_{c,t+1} \\
&\quad + \phi_d\sigma_{t+1}\eta_{d,t+2} + \phi_d\sigma_{t+2}\eta_{d,t+3} + \phi_d\sigma_t\eta_{d,t+1}
\end{aligned}$$

and the quarterly risk free rates is simply three times the current monthly risk free rate

$$\begin{aligned}
r^q_{f,t+3} &= 3 \cdot r_{f,t+3} \\
&= 3\zeta_\theta(z_{t+3}) + 3\phi_f\eta_{f,t+3}.
\end{aligned}$$

### B.2.1 Joint Distribution

We would like to describe the set containing the extended latent observations, the quarterly observations as the following

$$\begin{bmatrix} \Delta c_{t+2} \\ \Delta c_{t+3} \\ \Delta d_{t+2} \\ \Delta d_{t+3} \\ \Delta c^q_{t+3} \\ \Delta d^q_{t+3} \\ r^q_{m,t+3} \\ r^q_{f,t+3} \end{bmatrix} = \begin{bmatrix} \mu + x_{t+1} \\ \mu + x_{t+2} \\ \mu_d + \Phi x_{t+1} \\ \mu_d + \Phi x_{t+2} \\ 2\mu + x^q_{t+3} + \frac{2}{3}\Delta c_t + \frac{1}{3}\Delta c_{t-1} \\ 2\mu_d + \Phi x^q_{t+3} + \frac{2}{3}\Delta d_t + \frac{1}{3}\Delta d_{t-1} \\ \xi_\theta(z_{t+3}, z_{t+2}) + \xi_\theta(z_{t+2}, z_{t+1}) + \xi_\theta(z_{t+1}, z_t) + 3\mu_d + \Phi(x_{t+2} + x_{t+1} + x_t) \\ 3\zeta_\theta(z_{t+3}) \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & \sigma_{t+1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{t+2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_{dc}\sigma_{t+1} & 0 & 0 & \phi_d\sigma_{t+1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_{dc}\sigma_{t+1} & 0 & 0 & \phi_d\sigma_{t+2} & 0 & 0 & 0 & 0 \\ \sigma_t & \frac{2}{3}\sigma_{t+1} & \frac{1}{3}\sigma_{t+2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_{dc}\sigma_t & \frac{2}{3}\phi_{dc}\sigma_{t+1} & \frac{1}{3}\phi_{dc}\sigma_{t+2} & \phi_d\sigma_t & \frac{2}{3}\phi_d\sigma_{t+1} & \frac{1}{3}\phi_d\sigma_{t+2} & 0 & 0 & 0 & 0 \\ \phi_{dc}\sigma_t & \phi_{dc}\sigma_{t+1} & \phi_{dc}\sigma_{t+2} & \phi_d\sigma_t & \phi_d\sigma_{t+1} & \phi_d\sigma_{t+2} & \phi_m & \phi_m & \phi_m & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3\phi_f \end{bmatrix} \begin{bmatrix} \eta_{c,t+1} \\ \eta_{c,t+2} \\ \eta_{c,t+3} \\ \eta_{d,t+1} \\ \eta_{d,t+2} \\ \eta_{d,t+3} \\ \eta_{m,t+1} \\ \eta_{m,t+2} \\ \eta_{m,t+3} \\ \eta_{f,t+3} \end{bmatrix}, \tag{B.1}$$

where the noise terms are i.i.d Gaussian noises and we have

$$\begin{aligned} y^{(joint_Q)}_{t+3} &= \left\{ \Delta c_{t+2}, \Delta c_{t+3}, \Delta d_{t+2}, \Delta d_{t+3}, \Delta c^q_{t+3}, \Delta d^q_{t+3}, r^q_{m,t+3}, r^q_{f,t+3} \right\} \\ &= \mu^{(joint_Q)}_{t+3} + L^{(joint_Q)}_{t+3} \eta^{(joint_Q)}_{t+3} \end{aligned}$$

where $L^{(joint_Q)}_{t+3}$ is the corresponding loading matrix.

### B.2.2 Measurement Density

In order to calculate the measurement density of the quarterly observations and the inter-correlated market return, we take a part of the system of equations in B.1

$$y^Q_{t+3} = \left\{ \Delta c^q_{t+3}, \Delta d^q_{t+3}, r^q_{m,t+3}, r^q_{f,t+3} \right\} = \mu^{(observed)}_{t+3} + L^{(observed)}_{t+3} \eta^{(observed)}_{t+3}$$

where

$$\mu_{t+3}^{(observed)} = \mu_{t+3,5:8}^{(joint_Q)}, \quad L_{t+3}^{(observed)} = \{L_{t+3,i,j}^{(joint_Q)} : 5 \leq i, 1 \leq j\} \text{ and } \quad \eta_{t+3}^{(observed)} = \eta_{t+3,5:8}^{(joint_Q)}.$$

We obtain the full observation density

$$g(y_{t+3}^{Q}|z_t^q, z_t, z_{t+1}, z_{t+2}, z_{t+3}) = N\Big(\mu_{t+3}^{(observed)}, \Sigma_{t+3}^{(observed)}\Big)$$

where the covariance is $\Sigma_{t+3}^{(observed)} = L_{t+3}^{(observed)}(L_{t+3}^{(observed)})^T$. We want to sample the unobserved monthly variables $z_{t+3}^q = \{\Delta c_{t+3}, \Delta c_{t+2}, \Delta d_{t+3}, \Delta d_{t+2}\}$, based on the quarterly observations. We split the observation and their corresponding means into two parts, where the observation part $Y_2$ is known:

$$Y_1 = \begin{bmatrix} \Delta c_{t+2} \\ \Delta c_{t+3} \\ \Delta d_{t+2} \\ \Delta d_{t+3} \end{bmatrix}, Y_2 = \begin{bmatrix} \Delta c_{t+3}^q \\ \Delta d_{t+3}^q \\ r_{m,t+3}^q \end{bmatrix} \text{ and } \mu_1 = \mu_{t+3,1:4}^{(joint_Q)}, \mu_2 = \mu_{t+3,5:7}^{(joint_Q)},$$

and their corresponding covariance is splitted into four parts

$$\begin{aligned} \Sigma_{11} &= \{\Sigma_{t+3,i,j}^{(joint_Q)} : 1 \leq i \leq 4, 1 \leq j \leq 4\}, \\ \Sigma_{12} &= \{\Sigma_{t+3,i,j}^{(joint_Q)} : 1 \leq i \leq 4, 5 \leq j \leq 7\}, \\ \Sigma_{21} &= \{\Sigma_{t+3,i,j}^{(joint_Q)} : 5 \leq i \leq 7, 1 \leq j \leq 4\}, \\ Sigma_{22} &= \{\Sigma_{t+3,i,j}^{(joint_Q)} : 5 \leq i \leq 7, 5 \leq j \leq 7\}. \end{aligned}$$

If $Y_2$ is observed , the conditional distribution of $Y_1|Y_2$ is also normally distributed. The conditional mean and conditional covariance matrix are given by:

$$\begin{aligned} \mu^{(conditional)} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2) \\ \Sigma^{(conditional)} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Then we have

$$\hat{f}(\cdot|\Delta c_{t+3}^q, \Delta d_{t+3}^q, (z_t^q), z_t, z_{t+1}, z_{t+2}, z_{t+3}) = N(\mu^{(conditional)}, \Sigma^{(conditional)}).$$

## B.3 Sequential Monte Carlo for State Filtering and Likelihood Estimation.

The state-space model under consideration is both nonlinear and non-Gaussian, making it well-suited for state filtering via particle filters, especially when static parameters are known. The key concept is to represent the distributions of all random variables using a set of particles drawn directly from the state space and to approximate the posterior density $p(s_t|y_{1:t})$ by the empirical point-mass estimate $\hat{p}(s_t|y_{1:t})$.

Given $M$ samples $\{s_{t-1}^{(i)}, i = 1, 2, \ldots, M\}$ representing the filtering density $p(s_{t-1}|y_{1:t-1})$ at time $t-1$, the recursion

$$p(s_t|y_{1:t}) \propto \int p(y_t|s_t)p(s_t|s_{t-1})p(s_{t-1}|y_{1:t-1})ds_{t-1}$$

suggests an importance sampling strategy. First, sample from a known and easily accessible proposal density function, $q_t(s_t|s_{t-1})$, then assign importance weights $w_t$ to adjust for the difference between the target and the proposal,

$$w_t^{(i)} = \frac{p(y_t|s_t^{(i)})p(s_t^{(i)}|s_{t-1}^{(i)})}{q_t(s_t^{(i)}|s_{t-1}^{(i)})}. \tag{B.2}$$

The normalized weights are given by $W_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{M} w_t^{(j)}}$. To address particle degeneracy, focus on the most significant particles by resampling the population according to $W_t$. If the effective sample size $\text{ESS} = \frac{1}{\sum_{i=1}^{M}(W_t^{(i)})^2}$ falls below a certain threshold, resample the population to obtain an equally weighted sample representing the filtering distribution:

$$\hat{p}(s_t|y_{1:t}) = \frac{1}{M}\sum_{i=1}^{M} \delta_{s_t^{(a_t^i)}}(s_t).$$

Particle filters provide an estimate of the likelihood of observations as:

$$\hat{p}(y_{1:t}|\theta) = \prod_{l=2}^{t} \hat{p}(y_l|y_{1:l-1}, \theta)\hat{p}(y_1|\theta),$$

where

$$\hat{p}(y_l|y_{1:l-1},\theta) = \frac{1}{M}\sum_{i=1}^{M} w_l^{(i)}.$$

The likelihood estimate obtained through particle filters is unbiased, i.e., $E[\hat{p}(y_{1:t}|\theta)] = p(y_{1:t}|\theta)$, where the expectation is taken over all random variables used in the particle filter (Del Moral, 2004).

## B.4 Pseudo Code of S-UPF-CRN-SMC

### B.4.1 UKF-Based Proposal for Each Particle

Although the optimal proposal density in Doucet et al. (2000) is analytically intractable, we obtain a locally optimal approximation by running a Kalman filter on each particle for log-linearized components and an UKF for nonlinear components. The algorithmic steps are:

1. **Initialization** at $t = 0$: Draw $\{s_0^{(i)}\}_{i=1}^M$ from the initial distribution $p(s_0)$, assign each particle weight $1/M$.
2. **For time step** $t = 1, \ldots, T$ **and each particle** $i = 1, \ldots, M$**:**
   - Run the UKF on $s_{t-1}^{(i)}$ and $y_t$ to obtain mean $\bar{s}_t^{(i)}$ and variance $P_t^{(i)}$.
   - Sample
   $$s_t^{(i)} \sim \mathcal{N}\big(\bar{s}_t^{(i)}, P_t^{(i)}\big).$$
   - Update the weight for each particle using the importance weight formula (Equation B.2) and normalize.
   - Resample to obtain equally-weighted new particles $\{s_t^{(i)}\}_{i=1}^M$.

### B.4.2 S-UPF-CRN: Smoothed Unscented Particle Filter with CRNs

1. **Fix common randomness.** Draw and store a single random seed/vector $v$ to be reused in all Monte Carlo steps.

2. **Initialization ($t = 0$).**

   - Set UKF state mean $\mu_0$ and covariance $P_0$ for $(x, x_\lambda)$.
   - Generate $N$ bootstrap particles $\{\sigma_0^{(i)}\}_{i=1}^N$ using seed $v$.
   - Initialize the incremental likelihood $\hat{L} \leftarrow 1$.

3. **For each time $t = 1, \ldots, T$:**

   (a) *Bootstrap update for volatility with CRNs.* $\sigma_t^{(i)} \sim p(\sigma_t \mid \sigma_{t-1}^{(i)}; v)$ for $i = 1, \ldots, N$.

   (b) *UKF-based proposal for $(x, x_\lambda)$ with CRNs.*

   $$(x_t^{(i)},\, x_{\lambda,t}^{(i)},\, P_t^{(i)}) = \text{UKF_proposal}\big(x_{t-1}^{(i)}, x_{\lambda,t-1}^{(i)}, \theta, y_t;\, v\big).$$

   Here the unscented sigma-points and any internal sampling (e.g. process-noise draws) reuse the same fixed seed $v$.

   (c) *Compute importance weights.*

   $$w_t^{(i)} \propto \frac{p(y_t \mid x_t^{(i)}, \sigma_t^{(i)})\; p(\sigma_t^{(i)} \mid \sigma_{t-1}^{(i)})}{q_{\text{UKF}}(x_t^{(i)}, x_{\lambda,t}^{(i)})}.$$

   (d) *Gaussian smoothing with CRNs.* From $\{w_t^{(i)}, x_t^{(i)}, x_{\lambda,t}^{(i)}, \sigma_t^{(i)}\}$, compute a single weighted Gaussian approximation (mean $\mu_t^{approx}$, covariance $\Sigma_t^{approx}$); any sampling from this Gaussian also uses seed $v$.

   (e) *Increment likelihood.*

   $$\hat{L} \times = \mathcal{N}\big(y_t;\, \text{obs_mean}(\mu_t),\, \Sigma_t^{approx}\big).$$

4. **Output.** Return $\hat{L} \approx \hat{p}(y_{1:T} \mid \theta, v)$.

### B.4.3 S-UPF-CRN-SMC: SMC Sampler over $\theta$ Using S-UPF-CRN

1. **Inputs.**

   - Data $y_{1:T}$, prior $p(\theta)$.
   - Number of particles $N$, tempering levels $\{\alpha_0 = 0 < \alpha_1 < \cdots < \alpha_L = 1\}$.

- MCMC proposal $q(\theta' \mid \theta)$, fixed CRN seed $v$.
- Subroutine S-UPF-CRN$(y_{1:T}, \theta, N, v)$.

2. **Initialization.**
$$\theta_i^{(0)} \sim p(\theta), \quad w_i^{(0)} \leftarrow \tfrac{1}{N}, \quad i = 1, \ldots, N.$$

3. **Tempering loop ($\ell = 1, \ldots, L$).**

   (a) $\Delta\alpha = \alpha_\ell - \alpha_{\ell-1}$.

   (b) *Weight update.* For each $i$:
$$\hat{L}_i \leftarrow \text{S-UPF-CRN}(y_{1:T}, \theta_i^{(\ell-1)}, N, v), \quad w_i^{(\ell)} \leftarrow w_i^{(\ell-1)}\, \hat{L}_i^{\Delta\alpha}.$$

   (c) *Normalize weights.* $w_i^{(\ell)} \leftarrow w_i^{(\ell)} / \sum_j w_j^{(\ell)}$.

   (d) *Resampling.* Compute ESS $= 1/\sum_i (w_i^{(\ell)})^2$. If below threshold, resample $\{\theta_i^{(\ell)}\}$ and reset $w_i^{(\ell)} \leftarrow 1/N$; otherwise set $\theta_i^{(\ell)} \leftarrow \theta_i^{(\ell-1)}$.

   (e) *MCMC rejuvenation.* For each $i$:
$$\begin{aligned} &\theta' \sim q(\cdot \mid \theta_i^{(\ell)}), \quad \hat{L}' \leftarrow \text{S-UPF-CRN}(y_{1:T}, \theta', N, v), \\ &\log a = \big[\alpha_\ell \log \hat{L}' + \log p(\theta')\big] - \big[\alpha_\ell \log \hat{L}_i + \log p(\theta_i^{(\ell)})\big], \\ &\text{accept with prob. } \min(1, e^{\log a}). \end{aligned}$$

4. **Output.** Weighted ensemble $\{(\theta_i^{(L)}, w_i^{(L)})\} \approx p(\theta \mid y_{1:T})$.

## B.5 Empirical Validation with Covid-19 Data

When including COVID-period data—characterized by extreme volatility and structural breaks—the inference problem becomes even more challenging. Figure B.1 shows the same log-likelihood analysis across UPF and S-UPF. The mean log-likelihood (top panel) reveals increased dispersion while the standard deviation (bottom panel) exceeds the acceptable pseudo-marginal threshold for all filters, even at the highest particle count ($M = 2^{14}$). In fact, the standard deviation is around 10 at $M = 2^{14}$ for both UPF and S-UPF, whereas excluding COVID both filters achieve the desired standard deviation close to one at $M = 2^{14}$.

These results confirm that pseudo-marginal inference is not only computationally expensive but also unreliable in the presence of highly informative or irregular data, further motivating our proposed alternative.

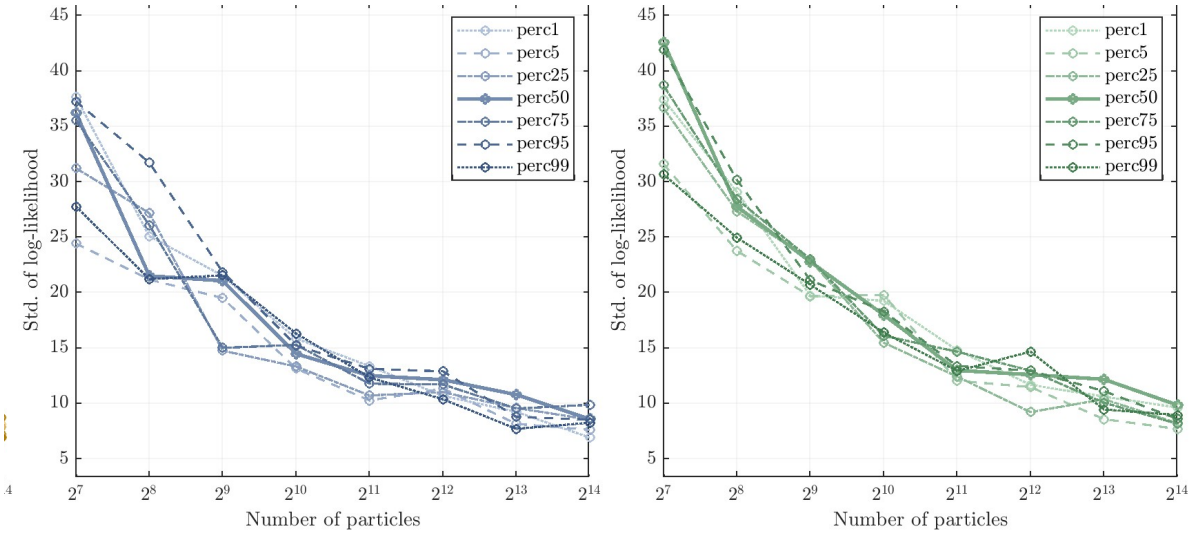



**Figure B.1:** Standard deviation of log-likelihood estimates for S-UPF (left) and UPF (right) across varying particle counts ($M = 2^7$ to $M = 2^{14}$). Results are averaged over 50 runs using 7 representative parameter sets.

## B.6 Additional Log-Likelihood Profiles

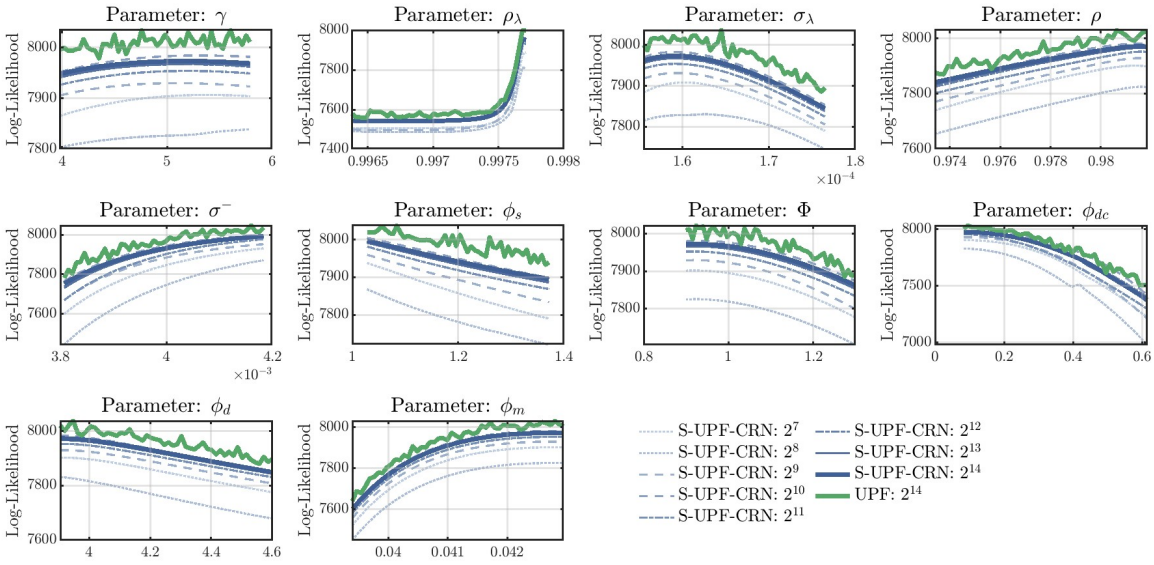



**Figure B.2:** Log-likelihood profile plots for the remaining parameter dimensions. The benchmark UPF (green) is compared to S-UPF-CRN with different particle counts (blue curves). This figure is a continuation of Figure 3.5.

## B.7 Additional Monte Carlo Simulation Results

This appendix presents the posterior density estimates for the nine model parameters not included in the main text analysis. These parameters primarily govern the long-run persistence ($\rho$), preference structure ($\gamma, \psi, \rho_\lambda$), and dividend dynamics ($\Phi, \phi_{dc}, \phi_d, \phi_s, \bar{\sigma}$).

Figure B.3 displays the comparison between the balanced-frequency (green) and mixed-frequency (blue) posteriors. Consistent with the findings in the main text, the inclusion of mixed-frequency data generally results in sharper posterior distributions (lower standard deviation) compared to the balanced-frequency benchmark.

However, these plots also provide further illustration of the bias-variance trade-off inherent in the S-UPF approximation. For parameters such as the leverage loading $\Phi$ (bottom left) and the risk aversion coefficient $\gamma$ (top left), the mixed-frequency posteriors are significantly tighter but exhibit a visible shift away from the true parameter value. By contrast, the balanced-frequency posteriors are centered more accurately on the truth

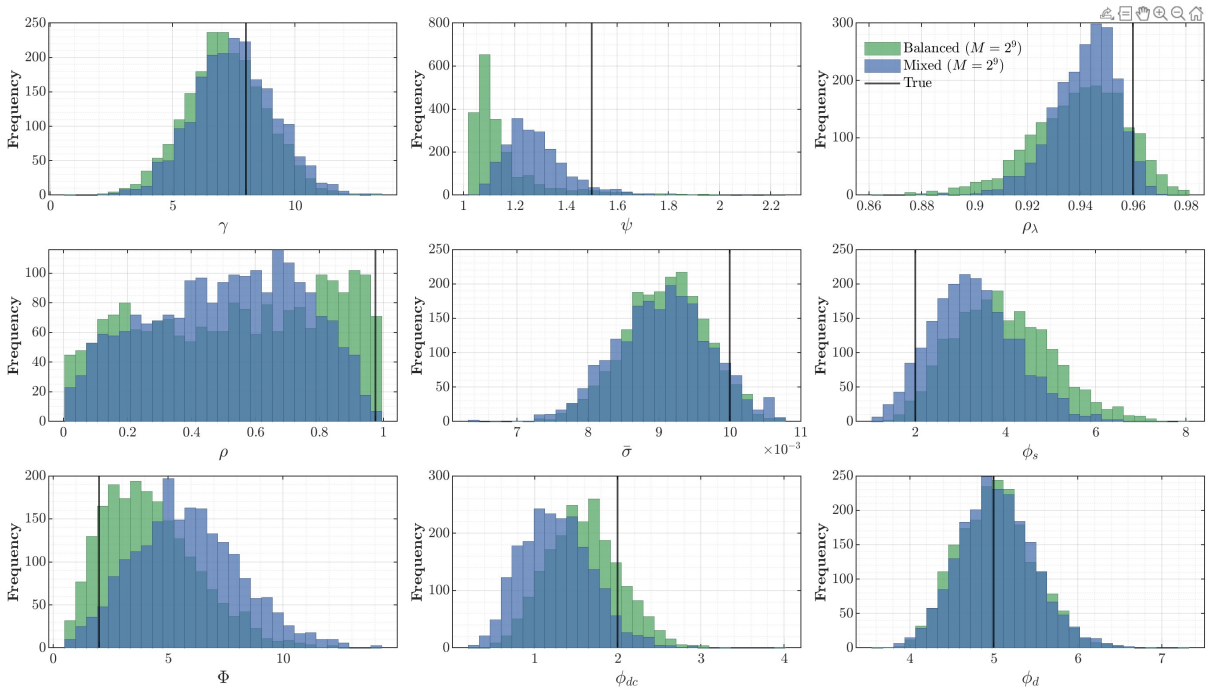



**Figure B.3:** Posterior density estimates for the remaining nine parameters. Green histograms represent the balanced-frequency scenario, blue histograms represent the mixed-frequency scenario, and black vertical lines indicate the true parameter values.

but are characterized by much higher dispersion. This confirms that while the Gaussian smoothing required for computational feasibility introduces a small structural bias, it succeeds in extracting information that reduces the overall parameter uncertainty.

# Appendix C

# Appendix to Chapter 4

## C.1 Density-tempered SMC

### C.1.1 Initialization

Start by allocating an initial probability $\bar{q}_i$ to each regressor where $i \in \mathbf{P}$. A straightforward and efficient approach would be setting $\bar{q}_i = Y_i^2 / \sum_{j=1}^{p} R_j^2$, with $R_i^2$ denoting the regression $R^2$ using a single Variable $i$. Let's equate $q_i^{p_s}(0)$ with $\bar{q}_i$ for all variables, and then use $q_i^{p_s}(0)$ to generate a sample $\mathbf{U}$ from $\mathbf{P}$, meaning, simulating $p_s$ regressors from the pool of $p$ regressors without replacement.

Because exactly $p_s$ regressors are needed, sampling can be performed sequentially. The first regressor is chosen from the $p$ potential regressors with a probability of $q_i^{p_s}(0)$ for $i \in \mathbf{P}$. Following this, we proceed to pick the second from the remaining $p-1$ regressors. Assuming the first one is variable $i$, the likelihood of selecting variable $j$ as the second is $\frac{q_i^{p_s}(0)}{1-q_i^{p_s}(0)}$ for $j \in \mathbf{P} \setminus \{i\}$. This reasoning is applied to the third regressor and continues until the final one in the permutation, i.e., $p_s$, with the probability of $q_k^{p_s}(0)$ for sampling variable $k$ from the leftover ones. If for some variable $i$, $q_i^{p_s}(0)$ equals zero, it will not be selected and can consequently be excluded from the initial set of regressors. Thus, $\mathbf{P}$ should be recognized as the set of regressors for which $q_i^{p_s}(0) > 0$. This permutation sampler, based on the distribution $q_i^{p_s}(0)$, is represented as $I(\mathbf{U} \in \mathbf{P}(p_s); q_i^{p_s}(0), i \in P)$.

Take note that the probability distribution of this sampler, being a product of the

mentioned probabilities, relies on the specific order of occurrence. In essence, the sampled points symbolize different permutations, which are simpler to sample and evaluate their probabilities. However, the regression solution is dependent only on the combination, meaning different permutations yielding the same combination will share an identical value. Consequently, the maximum is not unique over $\mathbf{U} \in \mathbf{p_s}$, but this is irrelevant to the solution.

### C.1.2 Reweighting and Resampling

We define the intermediate target, by advancing $\gamma$ from 0 to 1:

$$f_\gamma(\mathbf{U} \in P(p_s); \mathbf{R}, \mathbf{X}) \propto \Big( \frac{\hat{g}\big(\mathbf{R}, \hat{\theta}(\mathbf{X_U}); \mathbf{R}, \mathbf{X_U}\big)}{I(\mathbf{U} \in \mathbf{P}(p_s); q_i^{(p_s)}(0), i \in \mathbf{P})} \Big)^\gamma \cdot I(\mathbf{U} \in \mathbf{P}(p_s); q_i^{(p_s)}(0), i \in \mathbf{P})$$

It is obvious to see that $f_\gamma(\mathbf{U} \in P(p_s); \mathbf{R}, \mathbf{X})$ equals $I(\mathbf{U} \in \mathbf{P}(p_s); q_i^{(p_s)}(0), i \in \mathbf{P})$ for $\gamma = 0$, and equals to the target distribution (4.5), for $\gamma = 1$.

The distribution-tempered SMC technique moves the sample along a self-adapted control bridge by advancing $\gamma$ from 0 to 1. The self-adapted control rests with choosing so that the effective sample size (ESS) implied by the importance weight does not fall below a threshold $\rho M$ where $M$ is the intended size of the SMC sample and equals, say, 50%. Denote the incremental importance weight by $ESS = \frac{\left(\sum_{i=1}^M w_{\gamma,\gamma^{(j)}}(\mathbf{U}^{(i)})\right)^2}{\sum_{i=1}^M \left(w_{\gamma,\gamma^{(j)}}(\mathbf{U}^{(i)})\right)^2}$, where

$$w_{\gamma,\gamma^{(j)}}(\mathbf{U}^{(i)}) = \Big( \frac{\hat{g}\big(\hat{\theta}(\mathbf{R}, \mathbf{X_U}); \mathbf{R}, \mathbf{X_U}\big)}{I(\mathbf{U} \in \mathbf{P}(p_s); q_i^{(p_s)}(0), i \in \mathbf{P})} \Big)^{\gamma - \gamma^{(j)}}$$

By firstly setting $\gamma^{(0)} = 0$, we find a $\gamma^*$ such that the ESS is no less than $\rho M$. Note that this solution need not be exactly at $\rho M$, because it is just a control device to prevent the quality of sample from deteriorating too much. Use the incremental importance weight to resample in order to obtain an equally-weighted sample of $\mathbf{U}$, then set $\gamma^{(1)} = \gamma^*$.

### C.1.3 Support Boosting

Following resampling, the sample is likely to contain duplicate copies of certain U's, reflecting their higher importance weights and thus indicating a contraction of empirical support.

Before progressing $\gamma$ further, it's necessary to enhance this empirical support. This can be achieved by performing several Metropolis-Hastings (MH) moves until the cumulative realized acceptance rate hits a target, say, 100%. This is to ensure that empirical support has been adequately bolstered.

The SMC sample offers a natural basis to formulate a favorable proposal for executing MH moves. Compute $c_i^{(p_s)}(\gamma) = \sum_{j=1}^{M} \sum_{l=1}^{p_s} \chi_{\{\mathbf{U}_l^{(j)}=i\}}$, which is the total count of variable $i$ appearing in the sample of size $M$. Define a probability by $q_i^{(p_s)}(\gamma) = c_i^{(p_s)}(\gamma) / \sum_{j=1}^{p} c_i^{(p_s)}(\gamma)$, for $i = 1, 2, \ldots, p$, which reflects the relative importance of Variable i after the SMC algorithm has reached the stage indicated by $\gamma$. If $p_s$ regressors attain a probability of $1/p_s$, the sampler based on these probabilities will always generate different permutations of the same $p_s$ regressors. We use $\mathcal{Q}_{\mathbf{U}(\gamma)}$ to symbolize these probabilities derived from the sample. Instead of suggesting a new permutation in hopes of enhancing the solution, a more efficient and practical strategy is to replace a randomly selected subset of the existing permutation when $p_s$ is large. Notably, this deviates from the single best replacement algorithm of Soussen et al. (2015).

We use $h(\mathbf{U}^* \in \mathbf{P}(p_s) | \mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \mathcal{Q}_{\mathbf{U}(\gamma)})$ to denote the conditional distribution based on $\mathcal{Q}_{\mathbf{U}(\gamma)}$ for proposing regressors to replace $\tilde{A}$, a random subset of elements in permutation $\mathbf{U}$, where $\mathbf{U}_{-\tilde{A}}$ implies removing those elements in $\tilde{A}$ from $\mathbf{U}$. Given $\tilde{A}$, this probability is the standard permutation result for proposing element $i \in \mathbf{P} \setminus \mathbf{U}_{-\tilde{A}}$. Coupling this with the probability of sampling $\tilde{A}$ results in the overall proposal probability.

One could be tempted to propose a replacement subset similarly to the initialization step. However, if, due to some random chance, the current M $p_s$-variable combinations completely overlooked the variables in the optimal solution, the support boosting step would perpetually miss the optimal choice. A more dependable method of proposing new regressor permutations is thus to mix $h(\mathbf{U}^* \in \mathbf{P}(p_s) | \mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \mathcal{Q}_{\mathbf{U}(\gamma)})$ with $I(\mathbf{U}^* \in \mathbf{P}(p_s) | \mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \bar{q}_i, i \in \mathbf{P})$, the initialization sampler applying to $i \in \mathbf{P}/\mathbf{U}_{-\tilde{A}}$. Specifically, our proposal sampler is based on $h^{(w)}(\mathbf{U}^* \in \mathbf{P}(p_s) | \mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \mathcal{Q}_{\mathbf{U}(\gamma)}) = wh(\mathbf{U}^* \in \mathbf{P}(p_s) | \mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \mathcal{Q}_{\mathbf{U}(\gamma)} + (1-w) I(\mathbf{U}^* \in \mathbf{P}(p_s) | \mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \bar{q}_i, i \in \mathbf{P})$, where $w$ is set to 50% in our implementation.

The MH acceptance probability for replacing $\tilde{A}$, a random subset of elements in permu-

tation $U$, is

$$\begin{aligned}
&\alpha_{\gamma}^{(j)}\Big\{\mathbf{U}\in\mathbf{P}(p_s)\Rightarrow\mathbf{U}^*\in\mathbf{P}(p_s)\Big\}\\
&=\min\Big\{1,\frac{f_\gamma(\mathbf{U}^*\in P(p_s);\mathbf{R},\mathbf{X})}{f_\gamma(\mathbf{U}\in P(p_s);\mathbf{R},\mathbf{X})}\frac{h^{(w)}(\mathbf{U}\in\mathbf{P}(p_s)|\mathbf{U}^*_{-\tilde{A}}=\mathbf{U}_{-\tilde{A}};\mathcal{Q}_{\mathbf{U}(\gamma)})}{h^{(w)}(\mathbf{U}^*\in\mathbf{P}(p_s)|\mathbf{U}^*_{-\tilde{A}}=\mathbf{U}_{-\tilde{A}};\mathcal{Q}_{\mathbf{U}(\gamma)})}\Big\}
\end{aligned}\tag{C.1}$$

The acceptance probability establishes a Markov kernel, and by conventional reasoning, the target intermediate distribution in equation (4.5) is the stationary solution to this Markov kernel. In one cycle of support boosting, the MH step may substitute a variable number of elements to enhance combination diversity. For each cycle, an average acceptance rate can be calculated. The support boosting phase is deemed successfully concluded when the accumulated average acceptance rate across cycles has reached the target level, for which we set it at 500%.

Once the empirical support is adequately enhanced, the system is primed for progression from $\gamma = 0$ to eventually $\gamma = 1$. This advancement is achieved by continually executing the reweighting, resampling, and support boosting phases. Ultimately, the selected $p_s$ regressors correlate to the point in the final sample that offers the highest value of the distribution function. As previously noted, our sample points are permutations, and the regression solution depends solely on combinations. Therefore, there are likely numerous SMC sample points reaching the same highest value of the distribution function, but they all equate to the same choice of $p_s$ variables.

### C.1.4 Adding or Trimming Regressors

The ultimate SMC solution, represented by $\hat{\mathbf{U}}p_s$ with $p_s$ regressors, acts as a solid foundation for adding or removing variable(s) under a new target quantity $p'_s$, which is distinct from $p_s$. The modified initialization sampler applied here is a mixed sampler, designating a probability of $w'$ to an adding/removing operation. If $p'_s < p_s$, it's feasible to exclude $p_s - p'_s$ variables with equivalent probabilities from $\hat{\mathbf{U}}p_s$. If $p'_s > p_s$, we eliminate $p'_s - p_s$ variables from the remaining set of potential variables as usual to add to $\hat{\mathbf{U}}_{p_s}$, with the sampling executed according to the initial initialization probability as described before. We label the probability for this adding/removing sampler as $T(\mathbf{U}\in\mathbf{P}(p's);\hat{\mathbf{U}}p_s,\hat{q}_i,i\in\mathbf{P})$. The mixed sampler also assigns a probability of $1-w'$ to the original initialization sampling

distribution.

This modified initialization sampler for selecting $p'_s$ regressors can be represented by $I^{(w')}(\mathbf{U} \in \mathbf{P}(p_s)|\hat{\mathbf{U}}_{p_s}, \bar{q}_i, i \in \mathbf{P}) = w'T(\mathbf{U} \in \mathbf{P}(p'_s); \hat{\mathbf{U}}_{p_s}, \hat{q}_i, i \in \mathbf{P}) + (1 - w')I(\mathbf{U} \in \mathbf{P}(p_s)|\hat{\mathbf{U}}_{p_s}, \bar{q}_i, i \in \mathbf{P})$, where $w'$ is set to 0.1 in our implementation.

Accompanying this mixture initialization, we need to modify equation (4.5) to

$$f_\gamma(\mathbf{U} \in P(p'_s); \mathbf{R}, \mathbf{X}) \propto \Big( \frac{\hat{g}\big(\hat{\theta}(\mathbf{X}_\mathbf{U}); \mathbf{R}, \mathbf{X}_\mathbf{U}\big)}{I^{(w')}(\mathbf{U} \in \mathbf{P}(p'_s); \hat{\mathbf{U}}_{p_s} q_i^{(p_s)}(0), i \in \mathbf{P})} \Big)^\gamma I^{(w')}(\mathbf{U} \in \mathbf{P}(p'_s); \hat{\mathbf{U}}_{p_s} q_i^{(p_s)}(0), i \in \mathbf{P}) \tag{C.2}$$

Support boosting will be conducted per usual, but the MH acceptance probability must be altered accordingly to reflect the change to a mixture initialization sampler; that is,

$$\begin{aligned} & \alpha_\gamma^{(j)} \Big\{ \mathbf{U} \in \mathbf{P}(p'_s) \Rightarrow \mathbf{U}^* \in \mathbf{P}(p'_s) \Big\} \\ T \quad & = \min \Big\{ 1, \frac{f_\gamma(\mathbf{U}^* \in P(p_s); \mathbf{R}, \mathbf{X})}{f_\gamma(\mathbf{U} \in P(p_s); \mathbf{R}, \mathbf{X})} \frac{h^{(w,w')}(\mathbf{U} \in \mathbf{P}(p'_s)|\mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \mathcal{Q}_{\mathbf{U}(\gamma)})}{h^{(w,w')}(\mathbf{U}^* \in \mathbf{P}(p'_s)|\mathbf{U}^*_{-\tilde{A}} = \mathbf{U}_{-\tilde{A}}; \mathcal{Q}_{\mathbf{U}(\gamma)})} \Big\} \end{aligned} \tag{C.3}$$

where $h^{(w,w')}(\mathbf{U} \in \mathbf{P}(p'_s)|\cdot)$ is the same as $h^{(w)}(\mathbf{U} \in \mathbf{P}(p'_s)|\cdot)$ in equation (C.1) except that the initialization sampler is now a mixture sampler involving $w'$.

### C.1.5 Determining the Right Number of Regressors

In more straightforward terms, each time we increment our steps by one, we face the ongoing challenge of selecting an extra predictor from the remaining pool - the predictors not initially chosen by the SMC. Suppose we have already identified the $p_s$ optimal predictors, leaving us with the residual $n - p_s$ predictors. We aim to execute the SMC algorithm again, this time considering $p_s + 1$ predictors. Despite the initial classification of these $n - p_s$ predictors as insignificant, our task now requires us to pinpoint the most impactful within this subset. This scenario parallels a supremum statistic situation as we want to discover the most 'significant' predictor within an otherwise insignificant set. When we incorporate this extra predictor along with the already chosen $p_s$ predictors, we inherently conduct a multitude of hypothesis tests across the remaining pool of predictors. This necessitates a suitable adjustment of the significance level ($\alpha$-level) to account for the increased risk of Type

I errors that inherently arise from multiple hypothesis testing. This adjustment can be achieved through straightforward simulation, whereby we relax certain assumptions such as treating predictors as i.i.d. Furthermore, we disregard the potential discrepancy that the optimal subset of predictors for $p_s$ may not entirely overlap with the optimal subset of predictors for $p_s + 1$.

## C.2 Results of Unfiltered Data

In this appendix, we report results using the full unscreened design comprising 418 regressors. This comparative analysis yields a crucial methodological insight: while the SMC algorithm demonstrates robust convergence in high-dimensional, collinear environments, the absence of prescreening materially degrades interpretive clarity.

We observe that without prescreening, the selected predictor sets become heavily saturated with highly correlated variables, particularly regarding default probability horizons. As shown in Table C.1, the logit-heteroskedastic model simultaneously selects *Prior PD 1-Month*, *Prior PD 2-Week*, and various interactions thereof. Similarly, the beta model includes both *Prior PD 1-Week* and *Prior PD 1-Month* alongside multiple overlapping liquidity metrics. This simultaneous inclusion of multiple correlated horizons complicates economic inference; it becomes ambiguous whether the relevant signal stems from immediate distress (1-week) or intermediate deterioration (1-month). These findings suggest that light prescreening is not merely a tool for dimensional reduction, but a necessary condition for ensuring economic interpretability and avoiding signal redundancy.

**Table C.1:** Parameter Estimates for Logit-Heteroskedastic and Beta Models (Specification: Firm-Specific + Interactions, Without Prescreening)

**Logit-Heteroskedastic Model**

| Variable | Estimate | Std. Err. |
|---|---|---|
| **Mean** | | |
| *Domestic PD 3-Month* | -1.2779 | (0.1152) |
| *Yield Spread (7–10 year)* | -1.0066 | (0.1064) |
| *Distance to Default 1-Day* | 2.4099 | (0.0944) |
| *Global PD 3-Month * Utilities* | 2.6497 | (0.4023) |
| *Domestic PD 12-Month * Financials* | 2.8659 | (0.3588) |
| *Domestic PD 12-Month * Unsecured* | 3.0538 | (0.5859) |
| *US Yield Spread (5–7 year) * Financials* | -1.8815 | (0.2892) |
| *US Yield Spread (5–7 year) * Subordinated* | -1.2537 | (0.1830) |
| *Prior PD 1-Month * Unsecured* | -2.5719 | (0.4448) |
| *Prior PD 2-Week * Communications* | 1.0764 | (0.1387) |
| *Prior PD 1-Month * Convertible* | -0.6262 | (0.1747) |
| *Liquidity 1-Day * Real Estate* | -9.6274 | (0.4974) |
| *Liquidity 1-Day * Senior Unsecured* | -1.0233 | (0.1013) |
| **Variance** | | |
| *S&P500* | 0.9924 | (0.1526) |
| *Leverage Ratio 1-Day* | 0.2776 | (0.1327) |
| *S&P500 * Unsecured* | 2.3666 | (0.7108) |
| *Prior PD 1-Month * Health Care* | 4.0907 | (1.1488) |
| *Liquidity Ratio 1-Day * Real Estate* | -14.7110 | (1.2529) |

**Beta Model**

| Variable | Estimate | Std. Err. |
|---|---|---|
| **Mean** | | |
| *Domestic PD 3-Month* | -0.6165 | (0.0923) |
| *Yield Spread 7-10 Year* | -0.8348 | (0.0818) |
| *Distance to Default 1-Day* | 1.6550 | (0.0826) |
| *Global PD 3-Month * Utilities* | 2.1314 | (0.2728) |
| *Domestic PD 1-Month * Financials* | 0.8043 | (0.1587) |
| *Domestic PD 3-Month * Real Estate* | -1.5061 | (0.1166) |
| *Prior PD 1-Week * Senior Unsecured* | -0.4373 | (0.0901) |
| *Prior PD 1-Month * Communications* | 0.7901 | (0.1128) |
| *Leverage Ratio 1-Day * Subordinated* | -1.1879 | (0.1969) |
| *Liquidity Ratio 1-Day * Senior Unsecured* | -0.5939 | (0.0923) |
| **Dispersion (Precision)** | | |
| *Global PD 1-Month * Senior Unsecured* | -0.5055 | (0.1290) |
| *Yield Spread 7-10 Year * Convertible* | 0.5568 | (0.1660) |
| *Prior PD 1-Week * Real Estate* | -2.9593 | (0.3346) |
| *Distance to Default 1-Day * Utilities* | -0.9534 | (0.2845) |
| *Liquidity Ratio 1-Day * Senior Unsecured* | -0.2988 | (0.1222) |
| *Intercept* | -1.0033 | (0.0551) |

## C.3 Out-of-Sample Log-Likelihood Comparisons

While the main text focuses on $R^2$ and Conditional Kolmogorov (CK) tests as the primary measures of predictive and structural fit, this appendix provides the log-likelihood results for completeness.

**Table C.2:** In-sample (IS) and Out-of-sample (OOS) Log-Likelihoods (Normalized)

| **Model** | **Metric** | **Basic** | **Basic+Inter.(NF)** | **Basic+Firm** | **Basic+Firm+Inter.(All)** |
|---|---|---|---|---|---|
| Logit-Hetero. | IS log-lik. | 0.1712 | 0.2049 | 0.2231 | 0.3509 |
| | OOS log-lik. | 0.1628 | 0.1974 | 0.2027 | 0.3269 |
| Beta | IS log-lik. | 0.1488 | 0.1853 | 0.2778 | 0.3133 |
| | OOS log-lik. | 0.1477 | 0.1648 | 0.2459 | 0.2760 |

As observed, the logit-heteroskedastic model tends to achieve higher normalized log-likelihoods. However, the logit transformation ($\log(y/(1-y))$) significantly stretches the density near the boundaries of 0 and 1. This causes the maximum likelihood estimator to prioritize the fit of rare, extreme recovery events at the expense of the conditional mean. Because the beta model passes the CK test and provides superior $R^2$ on the original economic scale, it is preferred for modelling the expected recovery values, despite the likelihood values reported here.