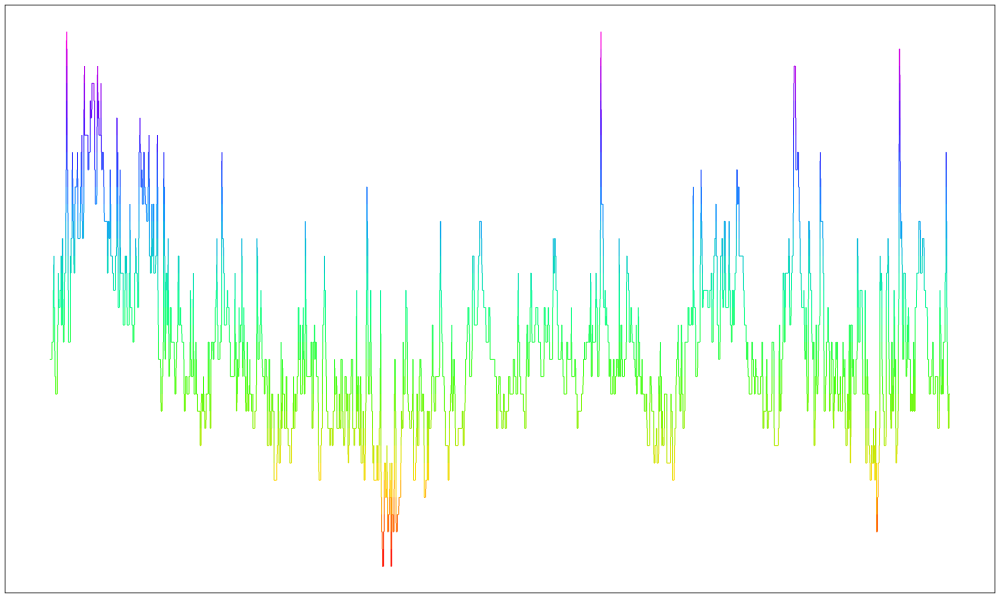
UNIVERSITY OF COPENHAGEN
DEPARTMENT OF MATHEMATICAL SCIENCES



Master's Thesis in Mathematical Finance

Youssef Mahmoud Raad
KU-id: zfw568

The Black-Scholes Option Pricing Model

A Markov-Switching Extension

Date: 22nd December 2025

Supervisor: Rolf Poulsen

Institute:	Institute for Mathematical Science
Department	Department of Mathematical Sciences
Author(s):	Youssef Mahmoud Raad
Title and subtitle:	The Black-Scholes Option Pricing Model: A Markov-Switching Extension
Description:	This Master's thesis explores extensions to the Black-Scholes option pricing model as a frequent critique of said model is it lack of adaptivity to economical turbulent environments. The extensions are via continuous and discrete state-space models often called Markov regime-switching models. The analysis is done using S&P 500 data (`^GSPC` and `^TR`).
Advisor:	Rolf Poulsen
Date:	22nd December 2025

Acknowledgements

A special thanks to Théo Michelot, Assistant Professor of Statistics at Dalhousie University, for his help throughout the process.
I would like to thank Rolf Poulsen, Professor of Mathematical Finance at the University of Copenhagen, for his supervision during the preparation project and this thesis. Thank you for the opportunity and for the chance to collaborate as a teaching assistant.
Above all and without comparison, my deepest thanks go to my mother. I would not have accomplished anything without her steady belief, patient encouragement and constant support.

This one is for you.

ةزيزعلا يمأ ىلإ

Note

We only provide proofs that are not widely known in both statistics and mathematical finance and/or that yield meaningful insights for learning. We will also include definitions when they are rarely encountered or when multiple competing formulations exist in the literature. The code is provided in Appendix A.1 as a `GitHub` repository link covering 55 files. Results and definitions used to establish the main results are collected in Appendix A.2.

Abstract

The Black-Scholes model (BSM) has long been a cornerstone of financial theory; however, its assumption of constant drift and volatility fails to capture the time-varying nature of asset returns, such as volatility clustering and abrupt regime shifts. This thesis investigates whether extending the BSM to include dynamic parameter evolution improves performance. Two extensions are proposed and implemented: a Black-Scholes Hidden Markov Model (BS-HMM) and a Black-Scholes Continuous State-Space Model (BS-SSM).

Using daily S&P 500 data from 1927 to 2025, the models are calibrated via Maximum Likelihood Estimation. In-sample analysis reveals that the extended models, particularly a 4-state BS-HMM and a factor-loaded continuous state-space model ($BS-SSM_{\beta}$), provide a superior fit to historical data compared to the static BSM. These models successfully identify distinct market phases, distinguishing between tranquil "bull" markets and high-volatility "crisis" regimes, such as the 1929 Crash and the 2008 Financial Crisis.

Despite the richer descriptive power and improved in-sample fit, out-of-sample evaluation on a hold-out period (2020-2025) indicates that the regime-switching extensions offer negligible gains in one-step-ahead point forecasting accuracy (MSE and RMSE) relative to the constant-parameter BSM. While the extended models produce more realistic, horizon-dependent forecast densities, the findings suggest that the added complexity of latent state inference does not translate into superior short-term predictive power for point forecasts.

CONTENTS

List of Symbols, Notation & Abbreviations

Symbol/Notation	Description
$\mathbb{N}$	Set of all positive integers
$\mathbb{R}$	Set of all real numbers
$\mathbb{P}$	Historical probability measure
$W_t^{\mathbb{P}}$	Brownian motion under a measure (here the measure is exemplified with $\mathbb{P}$)
Ω	Sample space
$\mathcal{F}$	Event space
$\{\mathcal{F}\}_{t\geq 0}$	Filtration
$(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, \mathbb{P})$	Filtered probability space
C_t	State occupied by Markov chain at time-t
p_i	Density function in state i
$\boldsymbol{P}(r)$	Diagonal matrix with ith diagonal element p_i
I_N	$N \times N$-dimensional diagonal matrix with i element 1 (identity matrix)
S_t	Random variable at time-t (asset price)
X_t	Random variable at time-t (log-return)
$\mathbf{c}^{(-t)}$	$(c_1, \ldots, c_{t-1}, c_{t+1}, \ldots, c_T)$ (and similarly for $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{V}$)
$\mathbf{C}^{(t)}$	$(C_1, C_2, \ldots, C_t)$ (and similarly for $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{V}$)
$\mathbf{C}_t^T$	$(C_t, C_{t+1}, \ldots, C_T)$ (and similarly for $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{V}$)
α_t	Forward probability, i.e. $\mathbb{P}(\mathbf{X}^{(t)} = \mathbf{x}^{(t)}, C_t = i)$
$\boldsymbol{\alpha_t}$	(Row) vector of forward probabilities
β_t	Backward probability, i.e. $\mathbb{P}(\mathbf{X}_{t+1}^T = \mathbf{x}_{t+1}^T \mid C_t = i)$
$\boldsymbol{\alpha}_t$	(Row) vector of forward probabilities
$\boldsymbol{\Gamma}$	Transition probability matrix of a Markov chain
γ_{ij}	(i,j)'th element in $\boldsymbol{\Gamma}$; probability of transitioning from state i to state j in a Markov chain
$\boldsymbol{\delta}$	Stationary distribution of a Markov chain
$\mathbf{1}_N$	N-dimensional vector of 1's
$\mathbf{0}_N$	N-dimensional row vector of 0's
$\mathbf{1}_{N\times N}$	$N \times N$-dimensional matrix filled with 1's
$\boldsymbol{e}_i$	$(0, \ldots, 0, 1, 0, \ldots, 0)$ i.e. a (row) vector of dimension T with a 1 in the t'th entry
T	Number of observations
N	Number of states
$\boldsymbol{\phi}$	Normalized vector of forward probabilities
$\boldsymbol{\psi}$	Predicted state probabilities
μ	Drift of the Black-Scholes model
σ	Volatility of the Black-Scholes model
ρ	Autoregressive parameter
σ_ε^2	variance of the innovations ε of the AR(1) process
Δ	Time-increment between some observations
$\mathcal{C}$	State space
H	Hessian matrix
SE$(\cdot)$	Standard Error of some estimator $\cdot$
$\mathcal{L}_T$	Likelihood function of T observations
ℓ_T	Log-likelihood function of T observations
$\xrightarrow{\mathcal{P}}$	Convergence in probability
$\xrightarrow{\mathcal{D}}$	Convergence in distribution
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{U}[a, b]$	Uniform distribution function over the range a to b
$\mathbb{1}_{\{A\}}$	Indicator function of some set A

Abbreviation	Description
BS	Black-Scholes
BSM	Black-Scholes model
HMM	Hidden Markov model (sometimes we use the combination BS-HMM)
SSM	State space model (sometimes we use the combination BS-SSM)
TR	Total Return
NTR	Net total return
AR(1)	Autoregressive process of order 1
t.p.m	Transition probability matrix
`nlm`	Non-linear maximization
ML/MLE	Maximum likelihood estimation
AIC	Akaike information criterion
BIC	Bayesian information criterion
SDE	Stochastic differential equation
a.s.	Almost surely
CI	Confidence interval
LLN	Law of large numbers
CLT	Central limit theorem
HAC	Heteroskedasticity- and autocorrelation-consistent
DAG	Directed acyclical graph

Introduction

The Black-Scholes model (BSM) of asset prices has long been a cornerstone of financial theory and practice. It models the asset price S_t as a geometric Brownian motion with constant expected rate of return μ and volatility σ, an assumption that leads to elegant analytical solutions for option pricing. Furthermore, the asset prices are log-normally distributed. This simplicity, however, comes at the cost of realism. In practice, asset returns exhibit time-varying volatility and occasional jumps or regime shifts that the log-normal Black–Scholes (BS) framework cannot capture. A proposed model to circumvent such jumps is Merton's Jump-Diffusion Model [37]. This model superimposes a jump component on a diffusion component of the asset price process. Formally, identifying jumps is challenging because large discrete returns can arise either from rare discontinuities or extreme diffusive shocks, making the two statistically indistinguishable in finite samples. Moreover, high-frequency data introduce microstructure noise and volatility clustering effects, which can mimic jumps and bias inference even in sophisticated econometric tests ([1], [3]). Indeed, [11] shows that jumps in financial asset prices are often erroneously identified and, in reality, are rare events that account for only a very small share of total price variation. Empirical returns often have heavier tails and more abrupt changes than the BSM assumes, indicating that the constant-σ assumption is too restrictive [45]. Indeed, it is often found that a single set of fixed parameters is inadequate to describe market dynamics across all periods. As market conditions evolve, the static BSM tends to lose predictive accuracy unless its parameters are frequently recalibrated [18]. This need for continual calibration of μ and σ underscores a key limitation of the classical model. That is, it is not well suited for forecasting in environments where volatility and other characteristics change over time.

To address the BSM's limitations in forecasting, this thesis considers two extensions that relax the assumption of constant parameters. The first is a BSM with a Hidden Markov Model (BS-HMM) for its parameters. In this regime-switching extension, the asset price still follows a diffusion as in BS, but its drift and/or volatility can switch between a finite set of states (regimes). These regime changes are governed by a hidden Markov chain. For instance, the market might alternate between "low-volatility" and "high-volatility" regimes, each with its own σ and possibly different μ, with probabilistic transitions between regimes. Such an HMM-based approach can capture phenomena like bull and bear market regimes or sudden volatility shifts that the classical model would miss. By allowing discrete shifts in parameters, the BS-HMM can dynamically adapt to structural changes in the data. We hypothesize that a BS-HMM will improve forecast accuracy by accounting for the possibility of regime shifts (e.g. an upcoming turbulent period) that a single-regime model would underestimate.

The second extension is a BSM with a Continuous State-Space Model (BS-SSM) for the parameters. In this approach, the drift and volatility are treated as latent continuous-time state variables that evolve according to their own stochastic process, specifically, an autoregressive pro-

cess of order 1. The observable data, the prices, are linked to these hidden states, forming a state-space model. Essentially, this is akin to a SVM. Volatility is no longer fixed, but changes over time in a continuous manner and we infer its path from the observed prices. Such state-space formulations are very flexible, allowing σ_t and μ_t to vary at each time step and capturing gradual shifts or cyclical patterns in volatility that a BS-HMM might not fully reflect. Conceptually, the BS-SSM treats the BS equation as having time-dependent parameters μ_t, σ_t governed by an underlying dynamical system. This continuous adaptation may better capture the nuanced evolution of market risk over time. Like the BS-HMM, the BS-SSM relaxes the constant parameter assumption, but it does so in a way that allows for more gradual, continuous changes rather than abrupt switches. We expect that by tracking a latent volatility factor, the BS-SSM can forecast future price distributions more accurately during periods of slowly changing market conditions or volatility trends.

The overarching research question addressed in this thesis is: Do these extended BSMs deliver superior out-of-sample forecasting performance compared to the BSM? In other words, we will test whether incorporating either discrete regime shifts or continuous stochastic parameter evolution leads to more accurate predictions of future asset prices than the traditional model with fixed μ and σ. This question is of both academic interest and practical importance. If the extended models can demonstrably improve forecast accuracy, it would suggest a pathway to better risk management and derivative pricing by acknowledging and modeling the non-stationarity in asset dynamics. Conversely, if the extensions do not improve forecasts, that finding is also informative as it would imply that the added complexity does not yields a sufficient performance gain for prediction and the basic BSM, perhaps with frequent recalibration, remains hard to beat.

To answer this question, we conduct an empirical comparison using historical data on the S&P 500 index. The methodology involves estimating each model (the classical BS, the BS-HMM and the BS-SSM) on a training sample and then evaluating their predictive performance on a hold-out sample. The models' parameters are fitted via maximum likelihood estimation, ensuring that each model is calibrated to the historical dynamics of the index. Once calibrated, each model generates out-of-sample forecasts of the S&P 500's price (or rather, log-return distribution) for the evaluation period. The key focus is on out-of-sample forecasting performance. By evaluating the predictions on data not used for estimation, we ensure a fair test of whether the additional flexibility of the HMM or state-space formulations translates into better predictive power. In short, we wish to extend the BSM by state-space models to examine if regime-switching does solve known issues of the BSM.

1 Data (I of II)

We briefly introduce the data and describe how dividends are computed. However, several key concepts have not yet been defined. We therefore split the data discussion into two non-consecutive sections.

The Standard and Poor's 500 (S&P 500) will be used throughout for analysis. The S&P 500 is a free-float–adjusted, value-weighted index of large-capitalization U.S. equities maintained by S&P Dow Jones Indices. By construction it spans major sectors of the U.S. economy and concentrates on firms with substantial public float, liquidity and operating history, yielding a diversified portfolio in which aggregate, rather than idiosyncratic, risk dominates variation. Its methodology and eligibility criteria (e.g., float adjustment, sector classification, profitability and size thresholds and scheduled re-constitutions) are transparent and stable over time, producing a well-documented data-generating mechanism that supports reproducible empirical work. We will work with the price version of the index, using daily data from Yahoo Finance (`^GSPC`). This series excludes cash dividends. S&P also publishes total return (TR) and net total return (NTR) variants (`^SP500TR`, `^SP500NTR`) that reinvest dividends gross or net of withholding taxes, respectively (see [52, 51, 50, 59]).
For asset-price analysis, the index offers several practical advantages. It captures a large share of total U.S. equity market capitalization and trading volume, which enhances the signal-to-noise ratio of return observations and reduces the impact of microstructure frictions at daily horizons. The series is long, clean and consistently adjusted for corporate actions, enabling inference across multiple macroeconomic episodes without survivorship bias tied to current constituents.

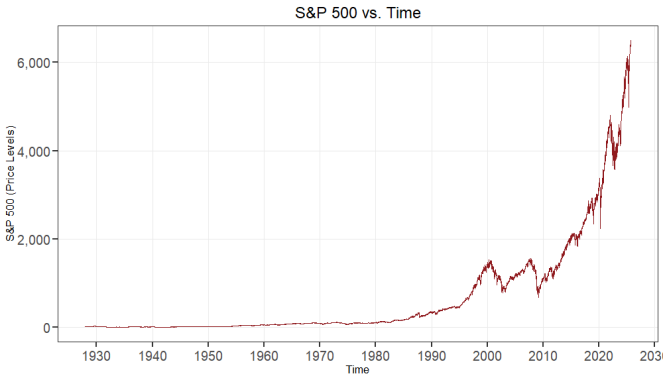


Figure 1: *S&P 500 index time series.*

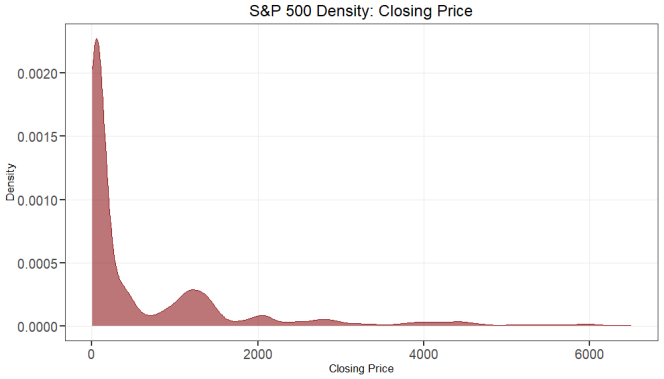


Figure 2: *Density of the S&P 500 index.*

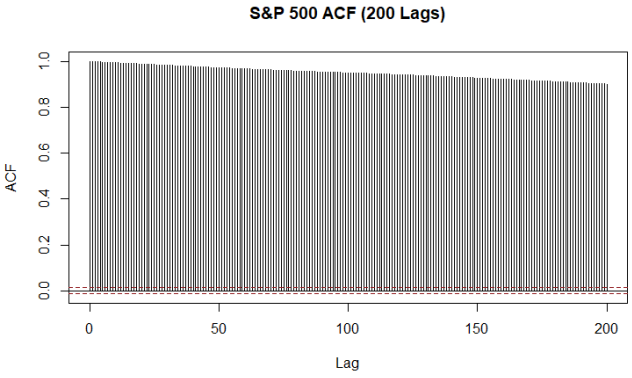


Figure 3: *Autocorrelation of closing prices (200 lags).*

The first observation in the dataset is recorded at market close on 1927-12-30 and the final observation is recorded at the end of the trading day on 2025-09-05. We estimate all models using data up to and including 2019-12-31 and we evaluate out-of-sample forecasts on the remaining observations.

The time series is displayed in Figure 1. The empirical distribution in Figure 2 concentrates substantial mass near the lower end of the support: the first quartile is 24.86, the median is 103.21 and the third quartile is 1076.22. The minimum and maximum observations are 4.40 and 6502.80, respectively. The density appears multimodal, with approximately six local maxima (the first three being the most pronounced). We therefore keep potential multimodality in mind when selecting the number of states. Finally, the autocorrelation function in Figure 3 exhibits strong persistence in levels for more than 200 lags, consistent with the nonstationarity of price levels.

Dividend Treatment and Construction of a Daily Dividend–Yield Series In this thesis, we adopt the Black–Scholes specification with constant parameters (μ, σ, q). With a continuous dividend yield q, the ex-dividend price index S_t satisfies

$$\frac{dS_t}{S_t} = (\mu - q)\,dt + \sigma\,dW_t^{\mathbb{P}}, \qquad d\ln S_t = \left(\mu - q - \tfrac{1}{2}\sigma^2\right)dt + \sigma\,dW_t^{\mathbb{P}},$$

so estimation on the price index identifies the capital-gains drift $\mu_{\text{cap}} := \mu - q$ rather than the total-return drift $\mu_{\text{tot}} := \mu_{\text{cap}} + q$.

Post-1988: Daily Estimation of q Let P_t denote the S&P 500 price index level (e.g., `^GSPC`) and T_t the corresponding total-return level (e.g., `^TR`). Over one trading day, for $t = 2, 3, \ldots, T$, define returns

$$r_t^{\text{PR}} = \frac{P_t}{P_{t-1}} - 1, \qquad r_t^{\text{TR}} = \frac{T_t}{T_{t-1}} - 1, \qquad 1 + r_t^{\text{TR}} = (1 + r_t^{\text{PR}})(1 + r_t^{\text{div}}).$$

Hence the dividend simple return is

$$r_t^{\text{div}} = \frac{1 + r_t^{\text{TR}}}{1 + r_t^{\text{PR}}} - 1,$$

and the log (continuous) dividend yield is

$$q_t^{(\log)} = \ln\left(\frac{T_t}{T_{t-1}}\right) - \ln\left(\frac{P_t}{P_{t-1}}\right).$$

Let $m_q := \mathbb{E}\left[q_t^{(\log)}\right]$ denote the daily mean log dividend yield and define the continuous annualized dividend yield as $q := 252 m_q$. We estimate m_q by the sample mean

$$\bar{q}^{(\log)} := \frac{1}{|\mathcal{T}|}\sum_{t \in \mathcal{T}} q_t^{(\log)}, \qquad \widehat{q} := 252\bar{q}^{(\log)},$$

where $\mathcal{T}$ denotes the full set of trading days in the sample.

Statistical inference for the constant-dividend estimator $\widehat{q}$ relies on heteroskedasticity- and

autocorrelation-consistent (HAC) standard errors of Newey-West type, with a naive i.i.d. standard error used as a benchmark. The construction of these standard errors and their extension to regime-specific dividend yields $\widehat{q}_i$, is detailed in Section 3.

Pre-TR Period: Shiller's Monthly Data to Daily Estimation of q Before daily TR series are readily available, we employ Shiller's long-run monthly S&P price and dividend data [48, 47] as a backfill. Let P_m be the month-end price index and D_m^{TTM} the trailing-twelve-month dividend total reported for month m; the implied monthly flow is $d_m := D_m^{TTM}/12$. The monthly log dividend yield is

$$q_m^{(\log)} = \ln\left(1 + d_m/P_m\right).$$

To obtain a daily series that preserves each month's total, distribute $q_m^{(\log)}$ evenly across the N_m business days in month m:

$$q_t^{(\log)} = \frac{q_m^{(\log)}}{N_m} \quad (t \in m),$$

so that $\sum_{t\in m} q_t^{(\log)} = q_m^{(\log)}$ by additivity of log-returns. This daily backfill is used only to compute the constant estimator $\widehat{q}$ over samples that include pre-1988 months. As a cross-check, S&P's Dividend Points indices track cumulative dividends in index points and reset annually [49]. Note that this method understates the variance of daily dividends in the pre-1988 sample compared to the post-1988 sample as it creates an artificial "step function" in the pre-1988 dividend volatility. We stress that q is treated as an exogenous variable estimated via sample means a posteriori to any model maximization.

2 Theory & Methodology

2.1 The Black–Scholes Model

Model under $\mathbb{P}$ The Black-Scholes model [5] (or Black-Scholes-Merton model [36]) assumes that there is a riskless asset with interest rate r such that the bank/money account B has time-varying dynamics

$$dB_t = rB_t dt, \qquad B_0 = 1,$$

and that the dynamics of the price of the underlying asset S under the historical probability measure $\mathbb{P}$ are

$$dS_t = (\mu - q)S_t dt + \sigma S_t dW_t^{\mathbb{P}}, \quad S_0 = s_0 > 0, \tag{1}$$

where $\mu \in \mathbb{R}$ is the expected return, $\sigma > 0$ the volatility and $q \geq 0$ denotes a constant dividend yield. For valuation purposes it is assumed that μ is an $\mathcal{F}$-adapted process.

The solution of Equation 1 can easily be found by using the transformation $Z_t = \log(S_t)$, where we assume that a solution exists and that S_t is a (strictly positive) solution. Itô's formula [4, Thm. 4.19] gives

$$\begin{aligned} dZ_t &= \frac{1}{S_t} dS_t + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)(dS_t)^2 \\ &= \frac{1}{S_t}\left((\mu - q)S_t dt + \sigma S_t dW_t^{\mathbb{P}}\right) + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)\sigma^2 S_t^2 dt \\ &= \left((\mu - q)dt + \sigma dW_t^{\mathbb{P}}\right) - \frac{1}{2}\sigma^2 dt. \end{aligned}$$

This leaves us with the equation

$$dZ_t = \left(\mu - q - \frac{\sigma^2}{2}\right) dt + \sigma dW_t^{\mathbb{P}}, \quad Z_0 = \log s_0. \tag{2}$$

Note that no counts of the r.v. Z_t is on the RHS of Equation 2. By implication, integrating yields

$$Z_t = \log(s_0) + \left(\mu - q - \frac{\sigma^2}{2}\right) t + \sigma W_t^{\mathbb{P}},$$

which means we obtain the solution to Equation 1 by reversing the log-transformation

$$S_t = s_0 \exp\left((\mu - q)t + \sigma W_t^{\mathbb{P}} - \frac{\sigma^2}{2} t\right). \tag{3}$$

From this point forward, we shall only consider a fixed time horizon $[0, \infty]$.

Distributional Properties By normality of the Wiener process increments [4, Def. 4.1], zero-mean property [4, Prop. 4.5] and Itô isometry [4, Prop. 4.5] the distribution of Z_t (and equivalently S_t) is

$$Z_t \sim \mathcal{N}\left(\log(s_0) + \left(\mu - q - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right)$$
$$\iff$$
$$S_t \sim \text{Lognormal}\left(\log(s_0) + \left(\mu - q - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right)$$

We proceed to work on an equidistant time grid $\{t_i = i\Delta,\ i = 0, \ldots, n\}$ over a fixed horizon T, where for $n \in \mathbb{N}$ $\Delta := T/n$ is the constant step size. For our purpose, $\Delta := 1/252$ corresponding to 252 business days in a year.
Define the continuously compounded return as

$$X_t := \log\left(\frac{S_t}{S_{t-1}}\right), \quad t = 2, 3, \ldots, T. \tag{4}$$

Using the solution, Equation 3 and the normality of the Wiener process increments [4, Def. 4.1] yields

$$X_t = \left(\mu - q - \frac{\sigma^2}{2}\right)\Delta + \sigma\left(W_t^{\mathbb{P}} - W_{t-1}^{\mathbb{P}}\right) \sim \mathcal{N}\left(\left(\mu - q - \frac{\sigma^2}{2}\right)\Delta, \sigma^2\Delta\right), \tag{5}$$

such that the corresponding probability density function for the r.v. X_t is

$$f_{X_t}(x_t) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left(-\frac{\left(x_t - \left(\mu - q - \frac{1}{2}\sigma^2\right)\Delta\right)^2}{2\sigma^2\Delta}\right), \quad x_t \in \mathbb{R}. \tag{6}$$

2.1.1 Likelihood Formulation & Parameter Estimation

Let $\{X_t\}_{t=1}^T$ denote T observed log-returns with distribution given in Equation 5. The BSM with parameter vector $\boldsymbol{\zeta} := (\mu, q, \sigma)$ is not identifiable in (μ, q), since only the difference $\mu - q$ enters the return distribution. Therefore, in the MLE we estimate the *capital-gains* drift μ_{cap} and recover the *total-return* drift via $\mu_{\text{total}} = \mu_{\text{cap}} + \widehat{q}$, where $\widehat{q}$ is the estimated continuous dividend yield (see Section 1 and Section 3). Henceforth, we set $\mu := \mu_{\text{cap}}$ unless stated otherwise. In practice, we find μ_{total} a posteriori, rather than a priori. The order does not change inference or estimation value.

For a realization $\mathbf{x}^{(T)} = (x_1, \ldots, x_T)$ of log-returns with step Δ, the likelihood and log-likelihood are

$$\mathcal{L}_T(\mu, \sigma) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\,\sigma^2\Delta}} \exp\left\{ -\frac{\left(x_t - \left(\mu - \frac{1}{2}\sigma^2\right)\Delta\right)^2}{2\,\sigma^2\Delta} \right\}$$

$$\iff$$

$$\ell_T(\mu, \sigma) = -\frac{T}{2}\log\left(2\pi\sigma^2\Delta\right) \; - \; \frac{1}{2\sigma^2\Delta}\sum_{t=1}^{T}\left(x_t - \left(\mu - \frac{1}{2}\sigma^2\right)\Delta\right)^2. \tag{7}$$

2.1.2 Simulation

To simulate the BSM (and the extended models), we need to develop a discretization scheme to simulate the continuous time process defined in Equation 1. To simulate paths for (S_t) at discrete times $\mathscr{T} = \{t_i\}_{i=1}^{T}$, we generate random samples of $(S_{t+\Delta})$ given (S_t) for some increment Δ. Repeatedly appending increments constructs the complete path $(S_t)_{t\in\mathscr{T}}$. We derive the Euler discretization scheme, often attributed to the work of [28].

Discretization Scheme Let $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space. Assume some r.v. X_t is driven by the SDE

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t^{\mathbb{P}}, \tag{8}$$

where $W_t^{\mathbb{P}}$ is a Wiener process under $\mathbb{P}$. Integrating Equation 8 from t to the incremented distance $t + \Delta$ yields

$$X_{t+\Delta} = X_t + \int_t^{t+\Delta} \mu(X_u, u)du + \int_t^{t+\Delta} \sigma(X_u, u)dW_u^{\mathbb{P}}. \tag{9}$$

At time-t, $\hat{X}_t$ is known. We aim to obtain the incremented, $\hat{X}_{t+\Delta}$. Euler scheme approximates the integrals using the left end-point rule, such that the deterministic integral of Equation 9 is approximated as the product of the integrand at time-t and the integration range Δ

$$\int_t^{t+\Delta} \mu(X_t, u)du \approx \mu(X_t, t)\int_t^{t+\Delta} du = \mu(X_t, t)\Delta.$$

Left end-points is a natural candidate as at time-t the value of $\mu(X_t, t)$ is known. Now, let $Z^{\mathbb{P}} \sim \mathcal{N}(0, 1)$. The stochastic integral is approximated as

$$\int_t^{t+\Delta} \sigma(X_u, u)dW_u^{\mathbb{P}} \approx \sigma(X_t, u)\int_t^{t+\Delta} dW_u^{\mathbb{P}} = \sigma(X_t, u)\left(W_{t+\Delta}^{\mathbb{P}} - W_t^{\mathbb{P}}\right) = \sigma(X_t, u)\sqrt{\Delta}Z^{\mathbb{P}},$$

because $W^{\mathbb{P}}_{t+\Delta} - W^{\mathbb{P}}_t$ and $\sqrt{\Delta}Z^{\mathbb{P}}$ are identically distributed [4, Def. 4.3]. Assembling the results yields the general form of the Euler discretization scheme:

$$\hat{X}_{t+\Delta} = \hat{X}_t + \mu(X_t, t)\Delta + \sigma(X_t, t)\sqrt{\Delta}Z^{\mathbb{P}}. \tag{10}$$

Applying Euler discretization to dS_t in Equation 1 by substituting the diffusion and drift of dS_t into Equation 10 yields the final discretization Euler scheme of the Black-Scholes' model dynamics

$$\hat{S}_{t+\Delta} = \hat{S}_t + \mu\hat{S}_t\Delta + \sigma\hat{S}_t\sqrt{\Delta}Z^{\mathbb{P}}. \tag{11}$$

The discretization scheme does induce a, varying but most often detrimental, discretization error to the continuous time process which is highly dependent on parameter subsets and discretization grid roughness as dictated by Δ. For a thorough examination and comments see the well-known work of [28, 7, 32].

Simulating We simulate the BSM with parameters $\mu = 0.05$, $\sigma = 0.15$, $S_0 = 100$ and $n = 25000$ over one realization of the stock price with daily observations, implying $\Delta = 1/252$ corresponding to approximately $25000/252 \approx 99$ years. Furthermore, we use the log-return X_t formulation in the implementation as given in Equation 4. The simulated stock price S_t is seen in Figure 4 where the used method of discretization was that developed in Section 2.1.2. The estimated values along side the true values are seen in Table 1. Parameter estimates were found using the `nlm`-function (Non-Linear Minimization) [13] in `R`. `nlm` is extremely popular for HMMs (see [56, 39, 34, 42, 60]) and provide (approximate) Hessians. Furthermore, `nlm` provides fast function evaluations via a Newtonian-style algorithm, as compared to i.e. the Nelder–Mead technique which is a heuristic search method.
As is quite evident, noting the Euler discretization scheme error, the estimates are accurate.

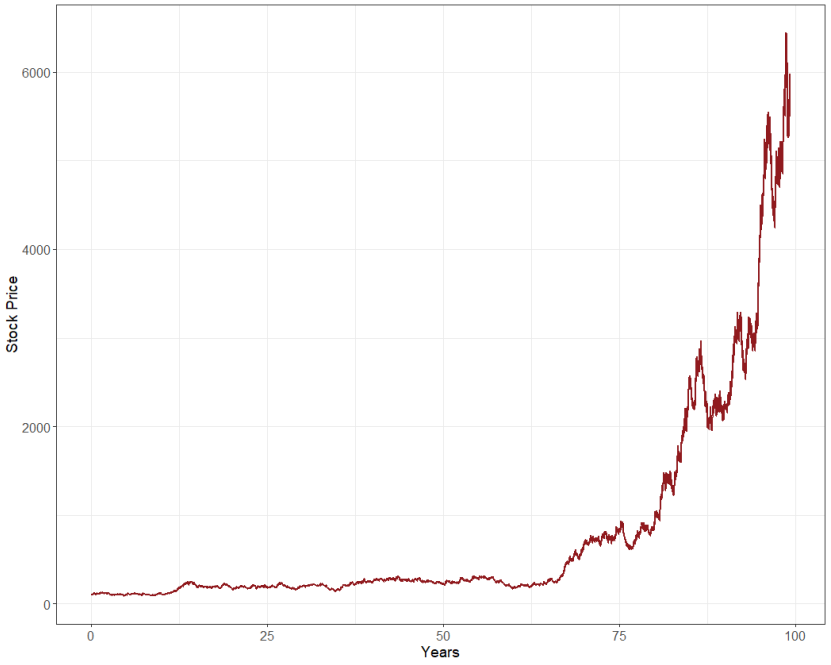


Figure 4: *A simulated stock price path S_t in the BSM.*

Parameter	True Value	Estimated Value	Relative Error (%)
μ	0.05000	0.05224	4.47
σ	0.15000	0.15002	0.0107

Table 1: *True vs. ML estimated BSM parameters, μ and σ. MLEs were found using direct numerical maximization using the `nlm`-function.*

2.2 Hidden Markov Models

2.2.1 Independent Mixture Models

A standard way to accommodate overdispersion is to use a finite mixture model, which represents unobserved heterogeneity by assuming the population comprises latent subgroups, each governed by its own distribution.

Consider an independent finite mixture with $N \in \mathbb{N} \setminus \{1\}$ components. Let $\delta_1, \ldots, \delta_N$ be mixing weights with $\delta_k \geq 0$ and $\sum_{k=1}^{N} \delta_k = 1$ and let $f_1, \ldots, f_N$ denote the component densities. Introduce a latent indicator $C \in \{1, \ldots, N\}$ with

$$\mathbb{P}(C = k) = \delta_k, \qquad k = 1, \ldots, N.$$

Conditional on $C = k$, the continuous random variable X has density $f_{X|C}(x \mid k) = f_k(x)$. The marginal density of X is therefore

$$f_X(x) = \sum_{k=1}^{N} \mathbb{P}(C = k)\, f_{X|C}(x \mid k) = \sum_{k=1}^{N} \delta_k f_k(x).$$

Figure 5 depicts the generative mechanism for $N = 2$. For each observation x_j, draw $C_j \in \{1, 2\}$ with $\mathbb{P}(C_j = i) = \delta_i$, then draw $x_j \sim f_{C_j}$. Across j, the indicators (C_j) are i.i.d. and conditional on (C_j), the observations (x_j) are independent.

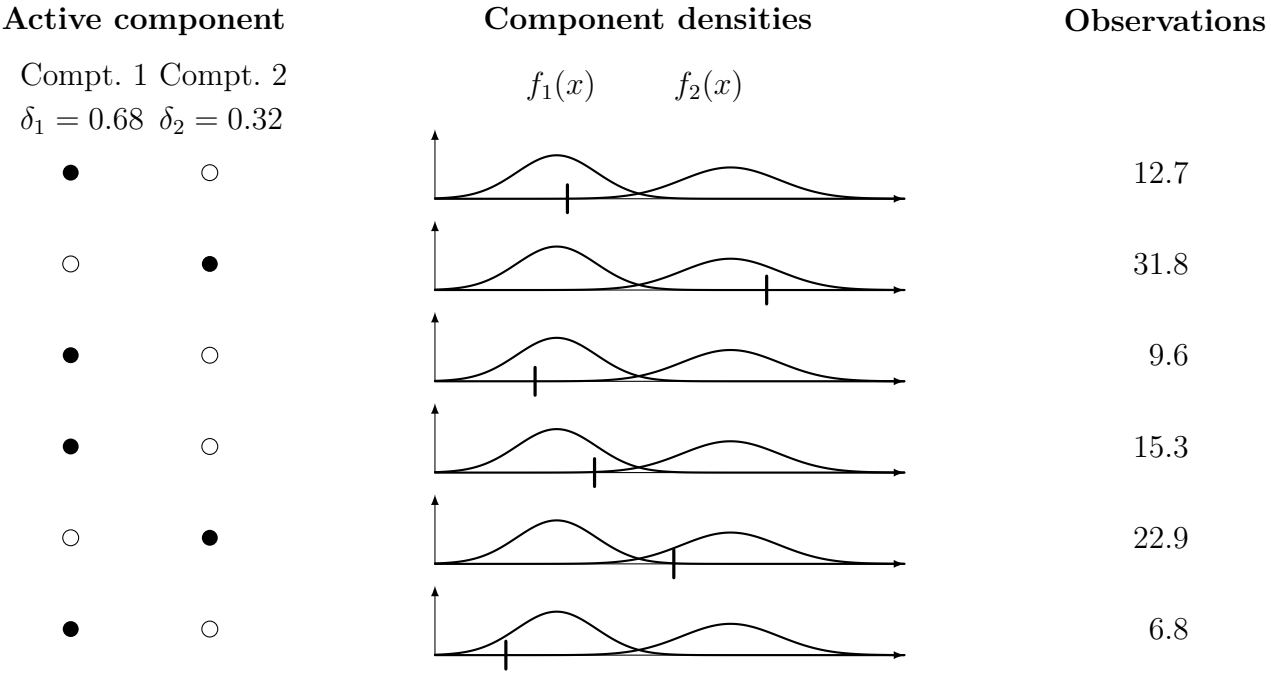


Figure 5: *A independent mixture model with 2 components, $\delta_1 = 0.68$ with corresponding density $f_1(x)$ and $\delta_2 = 0.32$ with corresponding density $f_2(x)$. The black veritcal line on the first axis represents the observation. A filled dot under a component denotes if it is active.*

Parameter Estimation Let $\boldsymbol{\zeta_1}, \ldots, \boldsymbol{\zeta_N}$ be the parameter vectors of the N-component distributions, $\delta_1, \ldots, \delta_N$ the mixing parameters where $\delta_k \geq 0$, $\sum_{k=1}^{N} \delta_k = 1$ and $x_1, \ldots, x_T$ be the T observations. The likelihood of a mixture model with N components is then given by

$$\mathcal{L}_T(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N, \delta_1, \ldots, \delta_N) = \prod_{j=1}^{T} \sum_{i=1}^{N} \delta_i f_{X_j,i}(x_j). \tag{12}$$

Thus, if the components are specified only by a single parameter, $2m - 1$ independent parameters have to be estimated by the component sum constraint.

Unbounded Likelihood in Mixtures The preceding theory for independent mixtures extends directly to a discrete setting by replacing densities with probability masses. A key difference is that the likelihood may become unbounded: in Gaussian mixtures, one can let a component mean coincide with an observation while its variance tends to zero, making the likelihood arbitrarily large. In such cases, it is often argued that the MLE does not exist [46, p. 4630].

A practical workaround is to use a discretized approximation of Equation 12,

$$\mathcal{L}_T^{\text{discrete}}(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N, \delta_1, \ldots, \delta_N) = \prod_{j=1}^{T} \sum_{i=1}^{N} \delta_i \int_{a_j}^{b_j} f_{X_j,i}(x)\, dx,$$

where (a_j, b_j) is the recording interval corresponding to the observed value x_j, i.e. the likelihood contribution is $\mathbb{P}(a_j < X_j < b_j)$. Alternatively, one may impose a strictly positive lower bound on component variances and maximise the likelihood subject to this constraint.

2.2.2 Markov Chains & Hidden Markov Models

Let $\{C_t\}_{t\in\mathbb{N}}$ be a sequence of discrete r.v.'s. $\{C_t\}_{t\in\mathbb{N}}$ is said to be a discrete-time Markov chain if, for all $t \in \mathbb{N}$ it satisfies the Markov property

$$\mathbb{P}(C_{t+1} \mid C_t, \ldots, C_1) = \mathbb{P}(C_{t+1} \mid C_t).$$

That is, conditional on the history up to and including time-t only depends on time-t. We will when possible use the notation $\mathbf{C}^{(t+1)} = (C_1, \ldots, C_t, C_{t+1})$ such that

$$\mathbb{P}(C_{t+1} = j \mid C_t = i, \ldots, C_1 = k) = \mathbb{P}(C_{t+1} = j \mid C_t = i),$$

where $j, i, k \in \mathcal{C}$ are states in the state space $\mathcal{C}$ at time $t = 1, 2, \ldots, T$. The Markov property is a relaxation of the assumption of independence and can be seen visualized in Figure 6.

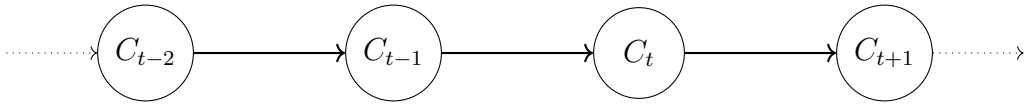


Figure 6: *A (first-order) Markov chain for the sequence of discrete r.v.'s* $\{C_t\}_{t\in\mathbb{N}}$.

An N-state hidden Markov model (HMM) $\{X_t\}_{t\in\mathbb{N}}$ is a dependent mixture model in which the conditional distribution of the observed process X_t is fully determined by an unobserved (latent) state variable $C_t \in \mathcal{C}$. The latent state process $\{C_t\}$ takes values in a finite discrete state space of size N and evolves as a discrete-time, time-homogeneous Markov chain. In particular, $\{C_t\}$ satisfies the Markov property and the observation process is conditionally independent of the past given the current state. In summary, the model assumptions are given by:

$$\mathbb{P}\left(C_t \mid \mathbf{C}^{(t-1)}\right) = \mathbb{P}\left(C_t \mid C_{t-1}\right), \quad t = 2, 3, \ldots,$$
$$\mathbb{P}\left(X_t \mid \mathbf{X}^{(t-1)}, \mathbf{C}^{(t)}\right) = \mathbb{P}\left(X_t \mid C_t\right), \quad t \in \mathbb{N}.$$

The model thus consists of an unobserved/hidden parameter process $\{C_t\}_{t\in\mathbb{N}}$ satisfying the Markov property and a state-dependent process $\{X_t\}_{t\in\mathbb{N}}$ in which the distribution of X_t depends exclusively on the time-t state, C_t.

The process is a noisy observation process in the sense that it is assumed to be produced by an underlying unobserved hidden state process, $\{C_t\}_{t\in\mathbb{N}}$. The distribution of X_t is conditionally independent of previous observations and all states except the current hidden state $i \in \mathcal{C}$:

$$f_{X_t|\mathbf{X}^{(t-1)},\mathbf{C}^{(t)}}(x_t \mid \mathbf{x}^{(t-1)}, \mathbf{c}^{(t)}) = f_{X_t|C_t}(x_t \mid i), \quad t = 2, 3, \ldots, T,$$

where f denotes a probability density function. The structure of a regular hidden Markov model can be seen in Figure 7, where the conditional independence can be intuitively understood.

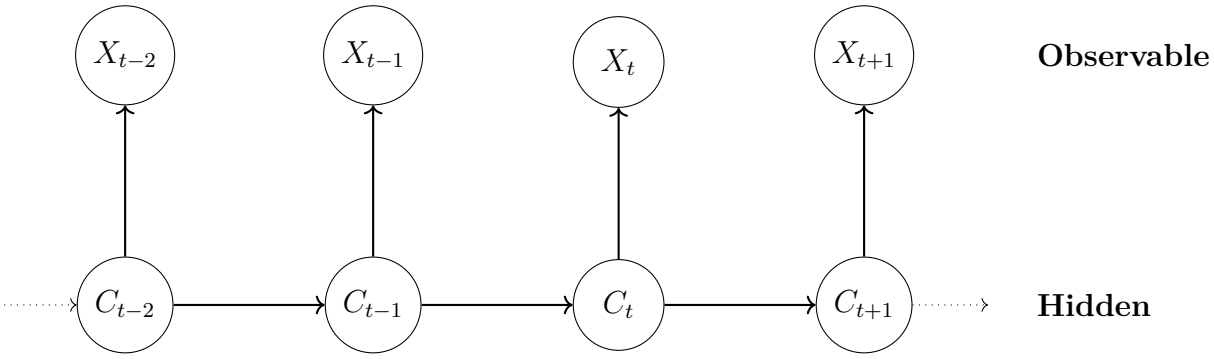


Figure 7: *A hidden Markov Model of order 1.*

The Markov chain induces dependence in the state-dependent process, meaning, the observations are independent of each other within states.

We will assume time-homogeneity of the Markov chain throughout this paper. The assumption of time-homogeneity of the Markov chain gives rise to the state transition probabilities, $\gamma \geq 0$, as

elements in the transition probability matrix (t.p.m.) $\Gamma \in \mathbb{R}^{N\times N}$ as

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} \end{bmatrix}, \quad \gamma_{ij} = \mathbb{P}(C_{t+1} = j \mid C_t = i) \in [0,1], \quad \sum_{j\in\mathcal{C}} \gamma_{ij} = 1. \tag{13}$$

For all $i, j \in \mathcal{C}$, let γ_{ij} denote the one-step transition probability from state i at time t to state j at time $t+1$, i.e., $\gamma_{ij} := \mathbb{P}(C_{t+1} = j \mid C_t = i)$. Under the assumption of time-homogeneity, these probabilities are invariant in t; equivalently, the transition law does not depend on the time index.

We define the concept of irreducibility.

Definition 2.1. *A Markov chain $\{C_t\}_{t=1,2,\ldots}$ with state space $\mathcal{C}$ is said to be irreducible if*

$$\forall i, j \in \mathcal{C}, \exists t \in \mathbb{N} : \mathbb{P}\left(C_{n+t} = j \mid C_n = i\right) > 0$$

for any valid time index n. Equivalently, every state can be reached from every other state with positive probability in some finite number of steps.

The concept of irreducibility can be seen intuitively in Figure 8.

(a) Irreducible **(b) Reducible: two classes** **(c) Reducible: absorbing**

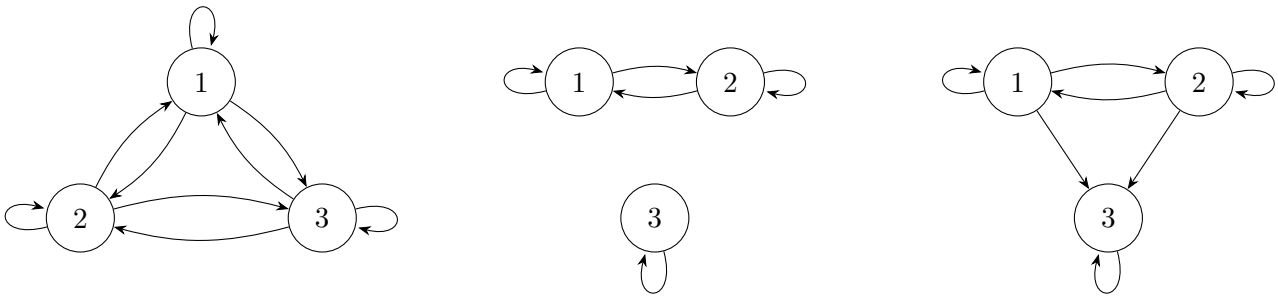


Figure 8: *Three finite Markov chains: (a) irreducible; (b) reducible with two communicating classes; (c) reducible with an absorbing class.*

Proceeding, the concept of a period and aperiodicity will be defined.

Definition 2.2. *For a state $i \in \mathcal{C}$, define its period as*

$$d(i) := \gcd\{\, t \geq 1 : (\mathbf{\Gamma}^t)_{ii} > 0 \,\}.$$

A state i is said to be aperiodic if $d(i) = 1$. A Markov chain is called aperiodic if all its states are aperiodic.

The concept of aperiodic is visualized in Figure 9.

(a) Periodic ($d = 2$) **(b) Aperiodic** ($d = 1$)

Figure 9: *Periodic vs aperiodic Markov chains. Adding a self-loop breaks periodicity.*

The unconditional probabilities of the state process refer to the probability of the process being in state i at time-t, unconditional of all previous states of the process. These are summarized in the row vector of probabilities

$$\boldsymbol{\delta}^{(t)} = \underbrace{\begin{bmatrix} \mathbb{P}(C_t = 1) & \cdots & \mathbb{P}(C_t = N) \end{bmatrix}}_{1 \times N}, \tag{14}$$

where the number of probabilities equals the number of states of the Markov chain. We let $\boldsymbol{\delta}^{(1)}$ denote the initial distribution of the Markov chain, which provides the probabilities of the process being in the different states at time-1. $\boldsymbol{\delta}^{(t)}$ allows for a convenient and surprisingly useful result.

Theorem 2.3. *Let $\boldsymbol{\delta}^{(t)}$ be defined as in Equation 14 and let $\Gamma = (\gamma_{ij})_{i,j=1}^{N}$ where $\gamma_{ij} = \mathbb{P}(C_{t+1} = j \mid C_t = i)$. Then, for $t \geq 1$,*

$$\boldsymbol{\delta}^{(t+1)} = \boldsymbol{\delta}^{(t)}\Gamma, \qquad \textit{and hence} \qquad \boldsymbol{\delta}^{(t)} = \boldsymbol{\delta}^{(1)}\Gamma^{t-1}.$$

Equivalently, $\boldsymbol{\delta}^{(t+1)} = \boldsymbol{\delta}^{(1)}\Gamma^{t}$.

Proof. Fix $j \in \{1, \ldots, N\}$. By the law of total probability,

$$\begin{aligned} \delta_j^{(t+1)} &= \mathbb{P}(C_{t+1} = j) = \sum_{i=1}^{N} \mathbb{P}(C_{t+1} = j \mid C_t = i)\, \mathbb{P}(C_t = i) \\ &= \sum_{i=1}^{N} \gamma_{ij}\, \delta_i^{(t)}. \end{aligned}$$

Writing this for all $j = 1, \ldots, N$ in vector form yields $\boldsymbol{\delta}^{(t+1)} = \boldsymbol{\delta}^{(t)}\Gamma$.

Iterating the recursion gives

$$\boldsymbol{\delta}^{(t)} = \boldsymbol{\delta}^{(t-1)}\Gamma = \cdots = \boldsymbol{\delta}^{(1)}\Gamma^{t-1},$$

which completes the proof. □

We now turn our attention to the stationary distribution. A Markov chain with a t.p.m. $\boldsymbol{\Gamma}$ is said

to have stationary distribution $\boldsymbol{\delta}$, a row vector with non-negative elements, if

$$\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}, \quad \boldsymbol{\delta}\mathbf{1}_N^\top = 1, \tag{15}$$

where $\mathbf{1}_N$ is a N-dimensional vector with entries 1. The first of the requirements in Equation 15 expresses the stationarity, i.e. moving forward in time is independent of the t.p.m., $\boldsymbol{\Gamma}$. The second is the requirement that $\boldsymbol{\delta}$ is indeed a probability distribution. To see how the stationarity and normalization conditions can be combined into a single linear system for $\boldsymbol{\delta}$, note that

$$\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta} \iff \boldsymbol{\delta} - \boldsymbol{\delta}\boldsymbol{\Gamma} = \mathbf{0}_N \iff \boldsymbol{\delta}(\boldsymbol{I}_N - \boldsymbol{\Gamma}) = \mathbf{0}_N,$$

where $\mathbf{0}_N$ is an N-dimensional row vector of zeros. Now, note that

$$\begin{aligned}
\sum_i \delta_i = 1 &\iff \begin{bmatrix} \delta_1 & \cdots & \delta_N \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = 1 \\
&\iff \begin{bmatrix} \delta_1 & \cdots & \delta_N \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \\
&\iff \boldsymbol{\delta}\mathbf{1}_{N\times N} = \mathbf{1}_N.
\end{aligned}$$

Adding the two equations, factoring out $\boldsymbol{\delta}$ and transposing, then yields the desired result

$$\begin{aligned}
\boldsymbol{\delta}(\boldsymbol{I}_N - \boldsymbol{\Gamma} + \mathbf{1}_{N\times N}) = \mathbf{1}_N &\iff (\boldsymbol{I}_N - \boldsymbol{\Gamma} + \mathbf{1}_{N\times N})^\top \boldsymbol{\delta}^\top = \mathbf{1}_N^\top \\
&\iff \left(\boldsymbol{I}_N - \boldsymbol{\Gamma} + \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \right)^\top \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_N \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}
\end{aligned}$$

Consequently, a Markov chain started from its stationary distribution will continue to have that distribution at all subsequent time points and we shall refer to such a process as a stationary Markov chain [60, p. 17]. Intuitively, a stationary distribution reflects the long-term proportion of time the model spends in each state.

To find the stationary distribution, one can obtain a explicit expression by solving the following system of equations [60, p. 18]

$$(\boldsymbol{I}_N - \boldsymbol{\Gamma} + \mathbf{1}_{N\times N})^\top \boldsymbol{\delta}^\top = \mathbf{1}_N^\top, \tag{16}$$

where $\boldsymbol{I}_N$ is a N-dimensional identity matrix, $\mathbf{1}_{N\times N}$ is a $N \times N$-dimensional matrix filled with 1's

and $\mathbf{1}_N$ is a vector filled with 1's.

Proposition 2.1. *Let $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times N}$ be row-stochastic, i.e. $\boldsymbol{\Gamma}\mathbf{1}_N = \mathbf{1}_N$ and define $\boldsymbol{J} := \mathbf{1}_N\mathbf{1}_N^\top$. A row vector $\boldsymbol{\delta} \in \mathbb{R}^{1\times N}$ satisfies*

$$\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}, \qquad \boldsymbol{\delta}\mathbf{1}_N = 1$$

if and only if

$$(\boldsymbol{I}_N - \boldsymbol{\Gamma} + \boldsymbol{J})^\top \boldsymbol{\delta}^\top = \mathbf{1}_N.$$

If $\boldsymbol{\Gamma}$ is irreducible, then $\boldsymbol{A} := (\boldsymbol{I}_N - \boldsymbol{\Gamma} + \boldsymbol{J})^\top$ is nonsingular and

$$\boldsymbol{\delta}^\top = \boldsymbol{A}^{-1}\mathbf{1}_N.$$

Proof. $(\Rightarrow)$ If $\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}$, then $\boldsymbol{\delta}(\boldsymbol{I}_N - \boldsymbol{\Gamma}) = \mathbf{0}_N$, hence

$$(\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top)\boldsymbol{\delta}^\top = \mathbf{0}_N.$$

If moreover $\boldsymbol{\delta}\mathbf{1}_N = 1$, then

$$\boldsymbol{J}\boldsymbol{\delta}^\top = \mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\delta}^\top = \mathbf{1}_N(\boldsymbol{\delta}\mathbf{1}_N) = \mathbf{1}_N.$$

Adding yields

$$(\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top + \boldsymbol{J})\boldsymbol{\delta}^\top = (\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top)\boldsymbol{\delta}^\top + \boldsymbol{J}\boldsymbol{\delta}^\top = \mathbf{1}_N,$$

which is equivalent to $(\boldsymbol{I}_N - \boldsymbol{\Gamma} + \boldsymbol{J})^\top\boldsymbol{\delta}^\top = \mathbf{1}_N$.

$(\Leftarrow)$ Assume $(\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top + \boldsymbol{J})\boldsymbol{\delta}^\top = \mathbf{1}_N$. Left-multiplying by $\mathbf{1}_N^\top$ and using $\mathbf{1}_N^\top\boldsymbol{\Gamma}^\top = \mathbf{1}_N^\top$ and $\mathbf{1}_N^\top\boldsymbol{J} = N\mathbf{1}_N^\top$ gives

$$N\,\mathbf{1}_N^\top\boldsymbol{\delta}^\top = \mathbf{1}_N^\top\mathbf{1}_N = N,$$

so $\boldsymbol{\delta}\mathbf{1}_N = 1$. Substituting back,

$$(\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top)\boldsymbol{\delta}^\top = \mathbf{1}_N - \boldsymbol{J}\boldsymbol{\delta}^\top = \mathbf{1}_N - \mathbf{1}_N = \mathbf{0}_N,$$

hence $\boldsymbol{\Gamma}^\top\boldsymbol{\delta}^\top = \boldsymbol{\delta}^\top$, i.e. $\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}$.

Now assume $\boldsymbol{\Gamma}$ is irreducible and let $\boldsymbol{x} \in \mathbb{R}^N$ satisfy $(\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top + \boldsymbol{J})\boldsymbol{x} = \mathbf{0}_N$. Left-multiplying by $\mathbf{1}_N^\top$ yields $N\,\mathbf{1}_N^\top\boldsymbol{x} = 0$, so $\mathbf{1}_N^\top\boldsymbol{x} = 0$. Then $\boldsymbol{J}\boldsymbol{x} = \mathbf{1}_N(\mathbf{1}_N^\top\boldsymbol{x}) = \mathbf{0}_N$ and the equation reduces to

$$(\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top)\boldsymbol{x} = \mathbf{0}_N \quad \Longrightarrow \quad \boldsymbol{\Gamma}^\top\boldsymbol{x} = \boldsymbol{x}.$$

For irreducible $\boldsymbol{\Gamma}$, the eigenvalue 1 of $\boldsymbol{\Gamma}^\top$ is simple, so the eigenspace $\{\boldsymbol{x} : \boldsymbol{\Gamma}^\top\boldsymbol{x} = \boldsymbol{x}\}$ is one-dimensional and spanned by a strictly positive vector $\boldsymbol{v}$. Thus any nonzero $\boldsymbol{x}$ with $\boldsymbol{\Gamma}^\top\boldsymbol{x} = \boldsymbol{x}$ has the form $\boldsymbol{x} = c\boldsymbol{v}$ and satisfies $\mathbf{1}_N^\top\boldsymbol{x} = c\,\mathbf{1}_N^\top\boldsymbol{v} \neq 0$, contradicting $\mathbf{1}_N^\top\boldsymbol{x} = 0$. Therefore $\boldsymbol{x} = \mathbf{0}_N$, so

$\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top + \boldsymbol{J}$ (and hence $\boldsymbol{A}$) is invertible and

$$\boldsymbol{\delta}^\top = (\boldsymbol{I}_N - \boldsymbol{\Gamma}^\top + \boldsymbol{J})^{-1}\mathbf{1}_N = \boldsymbol{A}^{-1}\mathbf{1}_N.$$

□

When transition probabilities are time-varying (e.g., as functions of covariates), a global stationary distribution does not exist [39, p. 14]. However, strictly conditioning on fixed covariate values allows for the derivation of a stationary distribution. In this work, we assume the Markov chain is stationary. This assumption is justified by the extensive duration of the observed data. Moreover, assuming stationarity significantly reduces the computational burden, as demonstrated in Proposition 2.2. Employment of Equation 16 throughout the code is used to estimate the stationary distribution to ease computational drag.

2.2.3 State-Dependent Distributions

The state-dependent distributions are the probability density functions of X_t given some state $i \in \mathcal{C}$ at time-t given by[1]

$$f_{i,X_t}(x_t) := f_{X_t|C_t}(x_t \mid i).$$

If the state process is stationary, the unconditional distribution of X_t can be given by

$$f_{X_t}(x_t) \overset{\dagger}{=} \sum_{i\in\mathcal{C}} f_{X_t,i}(x_t, i) = \sum_{i\in\mathcal{C}} f_{X_t|C_t}(x_t \mid i) f(C_t = i) = \sum_{i\in\mathcal{C}} \delta_i^{(t)} f_{i,X_t}(x_t) \overset{\dagger\dagger}{=} \sum_{i\in\mathcal{C}} \delta_i f_{i,X_t}(x_t), \quad (17)$$

where † follows from the law of total probability and †† by stationarity.

As the log-returns follow a normal distribution, we can directly specify the densities in Equation 17 by

$$f_{i,X_t}(x_t) = \frac{1}{\sqrt{2\pi\,\sigma_i^2\Delta}} \exp\left(-\frac{\left(x_t - (\mu_i - \frac{1}{2}\sigma_i^2)\Delta\right)^2}{2\,\sigma_i^2\Delta}\right), \quad x_t \in \mathbb{R}$$

Parameter Count As all the parameters are now defined for the BS-HMM we proceed to count the number of parameters to be estimated. The state process is characterized by $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$. The latter has $N \times (N-1) = N^2 - N$ free parameters due to the row-sum constraint (last equality of Equation 13). For a stationary Markov chain, we need not estimate the initial distribution as this equals the stationary distribution, which would otherwise yield N additional parameters

[1] The notation of the state-dependent density functions should not be confused with a joint density function. We will explicitly write i as a lower and first index when state-dependent densities are intended and not joint density functions.

(see Equation 14). As previously stated, we simply use Equation 16 to obtain the stationary distribution after estimating the transition probabilities.

Under conditional independence, the state-dependent process is governed by the state-dependent distributions. In the BSM setting these are parameterized by (μ, σ). If both are state-dependent, we require $2N$ state-parameters for estimation. The $N \times N$ transition probabilities is needed for the t.p.m., however, as the row-constraint states that the sum of transition probabilities in row i has to equal 1 this reduces to $N \times (N-1)$ parameters for estimation. In total we estimate

$$\#\text{Parameters}_2 = N^2 - N + 2N = N^2 + N.$$

Consider the case where exactly one of μ or σ is state-dependent (the other being globally state-independent). Simply replace the $2N$ by $N+1$, yielding

$$\#\text{Parameters}_1 = N^2 - N + (N+1) = N^2 + 1.$$

Lastly, if neither μ nor σ is state-dependent (both globally state-independent), we only need to estimate one count of both μ and σ, yielding

$$\#\text{Parameters}_0 = N^2 - N + 2.$$

For a visualization of the relation between state-dependent parameters and number of states see Figure 10.

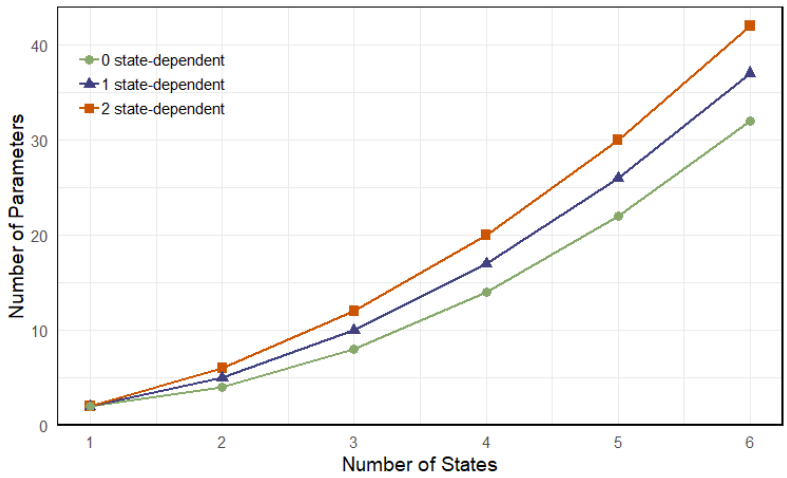


Figure 10: *Number of parameters as a function of the number of states N when 0, 1, or 2 of the BS parameters (μ, σ) are modeled as state-dependent.*

2.2.4 Likelihood Formulation & Parameter Estimation

The likelihood of a hidden Markov model has a convenient recursive form which is seen in the next result. Throughout, let the vector of parameters for estimation be denoted by $\boldsymbol{\zeta} = (\boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma})$.

Proposition 2.2. *Let $\{C_t\}_{t=1}^T$ be a time-homogeneous, finite N-state Markov chain on $\mathcal{C}$, with t.p.m. $\mathbf{\Gamma} = (\gamma_{ij})_{i,j=1}^N$. Consider a observation process $\{X_t\}_{t=1}^T$. The joint likelihood of observing $\{X_t\}_{t=1}^T$ is then given by*

$$\mathcal{L}_T(\boldsymbol{\zeta}) = \boldsymbol{\delta}^{(1)} \boldsymbol{P}(x_1) \mathbf{\Gamma} \boldsymbol{P}(x_2) \mathbf{\Gamma} \boldsymbol{P}(x_3) \cdots \mathbf{\Gamma} \boldsymbol{P}(x_t) \mathbf{1}^\top,$$

where $\boldsymbol{\delta}^{(1)}$ is the initial distribution, $\boldsymbol{P}(x)$ is the diagonal matrix with the state-dependent distribution $f_{1,X}(x)$, $f_{2,X}(x)$, …, $f_{N,X}(x)$ given in Equation 17 as elements and $\mathbf{\Gamma}$ is the t.p.m.. If $\boldsymbol{\delta}^{(1)}$ is the stationary distribution $\boldsymbol{\delta}$ of the Markov chain, then in addition

$$\mathcal{L}_T(\boldsymbol{\zeta}) = \boldsymbol{\delta} \mathbf{\Gamma} \boldsymbol{P}(x_1) \mathbf{\Gamma} \boldsymbol{P}(x_2) \mathbf{\Gamma} \boldsymbol{P}(x_3) \cdots \mathbf{\Gamma} \boldsymbol{P}(x_T) \mathbf{1}_N^\top.$$

´

Proof. Note that

$$\begin{aligned}
\mathcal{L}_T(\boldsymbol{\zeta}) &= f_{\mathbf{X}^{(T)}}\left(\mathbf{x}^{(T)}\right) \\
&= \sum_{c_1,\ldots,c_T=1}^{N} f_{\mathbf{X}^{(T)}|\mathbf{C}^{(T)}}\left(\mathbf{x}^{(T)} \mid \mathbf{c}^{(T)}\right) \mathbb{P}\left(\mathbf{C}^{(T)} = \mathbf{c}^{(T)}\right).
\end{aligned}$$

and by Lemma A.2.1

$$\begin{aligned}
\mathcal{L}_T(\boldsymbol{\zeta}) &= f_{\mathbf{X}^{(T)}}(\mathbf{x}^{(T)}) \\
&= \mathbb{P}(C_1) \prod_{k=2}^{T} \mathbb{P}(C_t \mid C_{t-1}) \prod_{k=1}^{T} f_{X_k|C_k}(x_k \,|\, C_k = c_k).
\end{aligned}$$

It then follows that

$$\begin{aligned}
\mathcal{L}_T(\boldsymbol{\zeta}) &= \sum_{c_1,c_2,\ldots,c_T=1}^{N} (\delta_{c_1} \gamma_{c_1,c_2} \gamma_{c_2,c_3} \cdots \gamma_{c_{T-1},c_T})(f_{c_1,X_1}(x_1) f_{c_2,X_2}(x_2) \ldots f_{c_T,X_T}(x_T)) \\
&= \sum_{c_1,c_2,\ldots,c_T=1}^{N} \delta_{c_1} f_{c_1,X_1}(x_1) \gamma_{c_1,c_2} f_{c_2,X_2}(x_2) \gamma_{c_2,c_3} \cdots \gamma_{c_{T-1},c_T} f_{c_T,X_T}(x_T) \\
&= \boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{\Gamma} \boldsymbol{P}(x_2) \ldots \mathbf{\Gamma} \boldsymbol{P}(x_T) \mathbf{1}_N^\top.
\end{aligned}$$

The last equality exploits the fact that a multiple sum of terms having a certain simple multiplicative form can in general be written as a matrix product (see [60, Ex. 7(b)] or Lemma A.2.5). If $\boldsymbol{\delta}$ is the stationary distribution of the Markov chain, we simply have

$$\boldsymbol{\delta} \boldsymbol{P}(x_1) = \boldsymbol{\delta} \mathbf{\Gamma} \boldsymbol{P}(x_1).$$

□

The recursive nature of the likelihood in Proposition 2.2 enables computationally efficient evaluation through numerical optimization. The likelihood is maximized using direct numerical methods, leveraging the forward probabilities and the forward algorithm.

Forward Probabilities The forward algorithm utilizes the forward probabilities which for $t = 1, 2, \ldots, T$ and $j \in \mathcal{C}$ are given as

$$\alpha_t(j) = f_{\mathbf{X}^{(t)}, C_t}\left(\mathbf{x}^{(t)}, j\right) = \mathbb{P}(C_t = j)\, f_{\mathbf{X}^{(t)} \mid C_t = j}\left(\mathbf{x}^{(t)}\right), \quad \boldsymbol{\alpha}_t = \left[\alpha_t(1) \ \ldots \ \alpha_t(N)\right]. \tag{18}$$

In other words, the forward probabilities contain information on the likelihood of the observations up to and including time-t. Also note that from the defintition of $\boldsymbol{\alpha}_t$ that, for $t = 1, 2, \ldots, T-1$, $\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \boldsymbol{P}(x_{t+1})$ which can be written in scalar form as

$$\alpha_{t+1}(j) = \left(\sum_{i \in \mathcal{C}} \alpha_t(i) \gamma_{ij} \right) f_{j, X_{t+1}}(x_{t+1}). \tag{19}$$

We are now able to prove the following result to justify their description as probabilities by utilizing the recursive form in Equation 19 and Lemma A.2.1.

Proposition 2.3. *For $t = 1, 2, \ldots, T$ and $j \in \mathcal{C}$,*

$$\alpha_t(j) = f_{\mathbf{X}^{(t)}, C_t}\big(\mathbf{x}^{(t)}, j\big).$$

Proof. Firstly, since $\boldsymbol{\alpha}_1 = \boldsymbol{\delta}\, \mathbf{P}(x_1)$ it follows that for $t = 1$

$$\alpha_1(j) = \delta_j f_{j, X_1}(x_1) = \mathbb{P}(C_1 = j)\, f_{X_1 \mid C_1}(x_1 \mid j) = f_{X_1, C_1}(x_1, j).$$

For some $t \in \mathbb{N}$ we then show it holds for $t + 1$:

$$\begin{aligned}
\alpha_{t+1}(j) &\overset{\dagger}{=} \sum_{i \in \mathcal{C}} \alpha_t(i)\, \gamma_{ij} f_{j, X_{t+1}}(x_{t+1}) \\
&= \sum_{i \in \mathcal{C}} f_{\mathbf{X}^{(t)}, C_t}\big(\mathbf{x}^{(t)}, i\big)\ \mathbb{P}(C_{t+1} = j \mid C_t = i)\ f_{X_{t+1} \mid C_{t+1}}(x_{t+1} \mid j) \\
&= \sum_{i \in \mathcal{C}} f_{\mathbf{X}^{(t)}, C_t, C_{t+1}}\big(\mathbf{x}^{(t)}, i, j\big)\ f_{X_{t+1} \mid C_{t+1}}(x_{t+1} \mid j) \\
&= \sum_{i \in \mathcal{C}} f_{\mathbf{X}^{(t+1)}, C_t, C_{t+1}}\big(\mathbf{x}^{(t+1)}, i, j\big) \\
&\overset{\dagger\dagger}{=} f_{\mathbf{X}^{(t+1)}, C_{t+1}}\big(\mathbf{x}^{(t+1)}, j\big),
\end{aligned}$$

where † is the scalar forward recursion (matrix–vector form $\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \mathbf{P}(x_{t+1})$) and we used the HMM conditional independences $X_{t+1} \perp (\mathbf{X}^{(t)}, C_t) \mid C_{t+1}$ and $\mathbf{X}^{(t)} \perp C_{t+1} \mid C_t$. Finally, †† is marginalization over the discrete r.v. C_t, that is, summing over i yields $f_{\mathbf{X}^{(t+1)}, C_{t+1}}(\mathbf{x}^{(t+1)}, j)$. □

Consequently, Equation 18 allows us to write the likelihood from Proposition 2.2 as

$$\mathcal{L}_t(\boldsymbol{\zeta}) = f_{\mathbf{X}^{(t)}}\left(\mathbf{x}^{(t)}\right) \overset{\dagger}{=} \sum_{j \in \mathcal{C}} f_{\mathbf{X}^{(t)}, C_t}\left(\mathbf{x}^{(t)}, j\right) = \sum_{j \in \mathcal{C}} \alpha_t(j).$$

where † follows from the law of total probability. The probability of the Markov chain occupying state-j at time-t given observations $\mathbf{X}^{(t)} = \mathbf{x}^{(t)}$, is its proportion of the forward probability at time-t for state j:

$$\mathbb{P}\left(C_t = j \mid \mathbf{X}^{(t)} = \mathbf{x}^{(t)}\right) = \frac{f_{C_t, \mathbf{X}^{(t)}}\left(j, \mathbf{x}^{(t)}\right)}{f_{\mathbf{X}^{(t)}}\left(\mathbf{x}^{(t)}\right)} = \frac{\alpha_t(j)}{\sum_{i \in \mathcal{C}} \alpha_t(i)}.$$

We can then state the (row) vector of forward probabilities for $t = 1, 2, \ldots, T$ as

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta} \boldsymbol{P}(x_1) \boldsymbol{\Gamma} \boldsymbol{P}(x_2) \cdots \boldsymbol{\Gamma} \boldsymbol{P}(x_t) = \boldsymbol{\delta} \boldsymbol{P}(x_1) \prod_{s=2}^{t} \boldsymbol{\Gamma} \boldsymbol{P}(x_s),$$

following the convention that an empty product is the identity matrix [60, p. 38]. Concluding, Proposition 2.2 states that $\mathcal{L}_T(\boldsymbol{\zeta}) = \boldsymbol{\alpha}_T \mathbf{1}_N^\top$. Furthermore, for $t \geq 2$ we defined $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t)$. This allows us to define the forward algorithm as:

$$\begin{aligned}
&\boldsymbol{\alpha}_1 = \boldsymbol{\delta} \boldsymbol{P}(x_1); \\
&\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t), \quad \text{for } t = 2, 3, \ldots, T; \\
&\mathcal{L}_T = \boldsymbol{\alpha}_T \mathbf{1}_N^\top.
\end{aligned}$$

Note, for a N-state HMM, $\boldsymbol{\delta}$ has N elements, $\boldsymbol{P}(x)$ has N elements (all in the diagonal) and $\boldsymbol{\Gamma}$ has $N \times N$ elements. For the forward algorithm, this implies that $\boldsymbol{\alpha}_t$ is a sum of N products consisting of a previous iteration, $\boldsymbol{\alpha}_{t-1}$, a transition probability γ_{ij} and a state-dependent probability $f_{i, X_t}(x_t)$, $i \in \mathcal{C}$. Hence, for each $t \in \{1, 2, \ldots, T\}$, there are N elements to be computed of $\boldsymbol{\alpha}_t$. Finally, this implies that the number of operations to calculate the likelihood of T observations is of order TN^2.

Backwards Probabilities Define

$$\boldsymbol{\beta}_t^\top = \boldsymbol{\Gamma}\boldsymbol{P}(x_{t+1})\boldsymbol{\Gamma}\boldsymbol{P}(x_{t+2})\cdots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}_N^\top = \left(\prod_{s=t+1}^{T}\boldsymbol{\Gamma}\boldsymbol{P}(x_s)\right)\mathbf{1}_N^\top,$$

with the convention that an empty product is the identity matrix. The case $t = T$ yields $\boldsymbol{\beta}_T = \mathbf{1}_N$. We then show that $\beta_t(j)$, the jth component of $\boldsymbol{\beta}_t$, can be identified as the the conditional density $f_{\mathbf{X}_{t+1}^T|C_t}\left(\mathbf{x}_{t+1}^T \mid i\right)$. It then immediately follows that for $t = 1, 2, \ldots, T$,

$$\begin{aligned}
\alpha_t(j)\beta_t(j) &= f_{\mathbf{X}^{(t)},C_t}\big(\mathbf{x}^{(t)}, j\big)\; f_{\mathbf{X}_{t+1}^T|C_t}\big(\mathbf{x}_{t+1}^T \mid j\big) \\
&\overset{(1)}{=} f_{\mathbf{X}^{(t)},C_t}\big(\mathbf{x}^{(t)}, j\big)\; f_{\mathbf{X}_{t+1}^T|\mathbf{X}^{(t)},C_t}\big(\mathbf{x}_{t+1}^T \mid \mathbf{x}^{(t)}, j\big) \\
&= f_{\mathbf{X}^{(t)},C_t}\big(\mathbf{x}^{(t)}, j\big)\; \frac{f_{\mathbf{X}^{(T)},C_t}\big(\mathbf{x}^{(T)}, j\big)}{f_{\mathbf{X}^{(t)},C_t}\big(\mathbf{x}^{(t)}, j\big)} \\
&= f_{\mathbf{X}^{(T)},C_t}\big(\mathbf{x}^{(T)}, j\big).
\end{aligned}$$

From the definition of $\boldsymbol{\beta}_t$ it follows that $\boldsymbol{\beta}_t^\top = \boldsymbol{\Gamma}\boldsymbol{P}(x_{t+1})\boldsymbol{\beta}_{t+1}^\top$. We now identify the backward probabilities as probabilities.

Proposition 2.4. *Assume* $\mathbb{P}(C_t = i) > 0$. *For* $t = 1, 2, \ldots, T-1$ *and* $i \in \mathcal{C}$,

$$\beta_t(i) = f_{\mathbf{X}_{t+1}^T,C_t}\left(\mathbf{x}_{t+1}^T \mid i\right).$$

Proof. We proof the proposition by induction. For $t = T - 1$

$$\beta_{T-1}(i) = \sum_j \mathbb{P}(C_T = j \mid C_{T-1} = i) f_{X_T,C_T}\left(x_T \mid j\right), \quad (\dagger)$$

since $\boldsymbol{\beta}_{T-1}^\top = \boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}_N^\top$. Furtermore, by Lemma A.2.3

$$\begin{aligned}
\mathbb{P}\left(C_T = j \mid C_{T-1=i}\right) f_{X_T|C_T}\left(x_T \mid j\right) &= \mathbb{P}(C_T = j \mid C_{T-1} = i) f_{X_t|C_{T-1},C_T}(x_T \mid i, j) \\
&= f_{X_T,C_{T-1},C_t}(x_T, i, j)/\mathbb{P}(C_{T-1} = i). \quad (\dagger\dagger)
\end{aligned}$$

Substituting $(\dagger\dagger)$ into $(\dagger)$ gives

$$\begin{aligned}
\beta_{T-1}(i) &= \frac{1}{\mathbb{P}(C_{T-1} = i)} \sum_j f_{X_T,C_{T-1},C_t}(x_T, i, j) \\
&= f_{X_T,C_{T-1}}(x_t, i)/\mathbb{P}(C_{T-1} = i) \\
&= f_{X_T|C_{T-1}}(x_t \mid i)
\end{aligned}$$

as required.

To demonstrate that validity at time $t+1$ implies validity at time t, we begin by observing that the recursive definition of β_t, in combination with the inductive hypothesis, gives

$$\beta_t(i) = \sum_j \gamma_{ij} f_{X_{t+1}|C_{t+1}}(x_{t+1}, j) f_{\mathbf{X}_{t+2}^T|C_{t+1}}(\mathbf{x}_{t+2}^T \mid j). \quad (\dagger\dagger\dagger)$$

However, Lemma A.2.2 and Lemma A.2.3 imply that

$$f_{X_{t+1}|C_{t+1}}(x_{t+1}, j) f_{\mathbf{X}_{t+2}^T}(\mathbf{x}_{t+2}^T \mid j) = f_{\mathbf{X}_{t+1}^T|C_t,C_{t+1}}(\mathbf{x}_{t+1}^T \mid i, j). \quad (\dagger\dagger\dagger\dagger)$$

Substitute from $(\dagger\dagger\dagger\dagger)$ into $(\dagger\dagger\dagger)$ which yields

$$\begin{aligned}
\beta_t(i) &= \sum_{j\in\mathcal{C}} \mathbb{P}(C_{t+1} = j \mid C_t = i) f_{\mathbf{X}_{t+1}^T|C_t,C_{t+1}}(\mathbf{x}_{t+1}^T \mid i, j) \\
&= \frac{1}{\mathbb{P}(C_t = 1)} \sum_{j\in\mathcal{C}} f_{\mathbf{X}_{t+1}^T,C_t,C_{t+1}}\left(\mathbf{x}_{t+1}^T, i, j\right) \\
&= \frac{f_{\mathbf{X}_{t+1}^T,C_t}\left(\mathbf{x}_{t+1}^T, i\right)}{\mathbb{P}(C_t = i)} \\
&= f_{\mathbf{X}_{t+1}^T|C_t}\left(\mathbf{x}_{t+1}^T \mid i\right)
\end{aligned}$$

which is the required conditional probability. □

Scaling the Likelihood Let $\mathcal{L}_t(\boldsymbol{\zeta})$ denote the likelihood of the observations up to time t, given a fixed parameter specification $\boldsymbol{\zeta}$ of a HMM. Under suitable regularity conditions, [30] established the existence of a constant $h \in \mathbb{R}$ such that the normalized log-likelihood converges almost surely:

$$\lim_{t\to\infty} \frac{1}{t} \log \mathcal{L}_t(\boldsymbol{\zeta}) = h, \quad \text{a.s.}$$

This result implies that asymptotically, $\mathcal{L}_t(\boldsymbol{\zeta}) \approx e^{th}$. Consequently, the likelihood exhibits exponential behavior as $t \to \infty$:

- if $h < 0$, $\mathcal{L}_t(\boldsymbol{\zeta})$ converges to 0 almost surely (exponential decay);
- if $h > 0$, $\mathcal{L}_t(\boldsymbol{\zeta})$ diverges to ∞ almost surely (exponential growth).

This exponential scaling presents a significant computational challenge. Direct evaluation of the likelihood will rapidly result in numerical underflow or overflow, necessitating the use of log-space calculations or scaling procedures.
As such, observe firstly from Proposition 2.2, that the HMM likelihood is a product of matrices and not scalars. Consequently, it is not possible to circumvent numerical underflow by computing the logarithm of the likelihood as the sum of logarithms of its factors.

As such, we adapt the method used by [60, p. 48] (although heavily inspired by [16, p. 78]): For $t = 1, \ldots, T$ define the standardised vector of forward probabilities at time-t as:

$$\boldsymbol{\phi}_t = \frac{\boldsymbol{\alpha}_t}{\boldsymbol{\alpha}_t \mathbf{1}_N^\top} = \frac{\boldsymbol{\alpha}_t}{\sum_{j \in \mathcal{C}} \alpha_t(j)}, \quad \boldsymbol{\phi}_t = [\phi_t(1) \ldots \phi_t(N)], \quad \sum_{j \in \mathcal{C}} \phi_t(j) = 1.$$

$\boldsymbol{\phi}_t$ are the normalized forward probabilities, which are far less susceptible to numerical underflow. For $t = 1$:

$$\boldsymbol{\phi}_1 = \frac{\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_1 \mathbf{1}_N^\top} = \frac{\boldsymbol{\delta}_0 \boldsymbol{P}(x_1)}{\boldsymbol{\delta}_0 \boldsymbol{P}(x_1) \mathbf{1}_N^\top}.$$

For $t = 2, \ldots, T$:

$$\boldsymbol{\phi}_t = \frac{\boldsymbol{\alpha}_t}{\boldsymbol{\alpha}_t \mathbf{1}_N^\top} = \frac{\boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t)}{\boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t) \mathbf{1}_N^\top} = \frac{\boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t) / (\boldsymbol{\alpha}_{t-1} \mathbf{1}_N^\top)}{\boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t) \mathbf{1}_N^\top / (\boldsymbol{\alpha}_{t-1} \mathbf{1}_N^\top)} = \frac{\boldsymbol{\phi}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t)}{\boldsymbol{\phi}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t) \mathbf{1}_N^\top}.$$

In other words, we have a algorithm which uses scalar multiplication as opposed to matrix multiplication. To see why this is actually the case, we derive the likelihood, $\mathcal{L}_T(\boldsymbol{\zeta})$, in terms of $\boldsymbol{\phi}$ as opposed to $\boldsymbol{\alpha}$.
Firstly, using $\boldsymbol{\alpha}_0 = \boldsymbol{\delta}$, note that

$$\mathcal{L}_T(\boldsymbol{\zeta}) = \boldsymbol{\alpha}_T \mathbf{1}^\top = \frac{\boldsymbol{\alpha}_1 \mathbf{1}^\top}{\boldsymbol{\alpha}_0 \mathbf{1}^\top} \frac{\boldsymbol{\alpha}_2 \mathbf{1}^\top}{\boldsymbol{\alpha}_1 \mathbf{1}^\top} \cdots \frac{\boldsymbol{\alpha}_T \mathbf{1}^\top}{\boldsymbol{\alpha}_{T-1} \mathbf{1}^\top} = \prod_{t=1}^{T} \frac{\boldsymbol{\alpha}_t \mathbf{1}^\top}{\boldsymbol{\alpha}_{t-1} \mathbf{1}^\top}, \tag{20}$$

where $\frac{\boldsymbol{\alpha}_t \mathbf{1}^\top}{\boldsymbol{\alpha}_{t-1} \mathbf{1}^\top} \in \mathbb{R}$. This allows us to find the log-likelihood function using Equation 20

$$\begin{aligned}
\ell_T(\boldsymbol{\zeta}) &= \log \mathcal{L}_T(\boldsymbol{\zeta}) \\
&= \log \prod_{t=1}^{T} \frac{\boldsymbol{\alpha}_t \mathbf{1}_N^\top}{\boldsymbol{\alpha}_{t-1} \mathbf{1}_N^\top} \\
&= \sum_{t=1}^{T} \log \left(\frac{\boldsymbol{\alpha}_t \mathbf{1}_N^\top}{\boldsymbol{\alpha}_{t-1} \mathbf{1}_N^\top} \right) \\
&= \log \left(\frac{\boldsymbol{\alpha}_1 \mathbf{1}_N^\top}{\boldsymbol{\alpha}_0 \mathbf{1}_N^\top} \right) + \sum_{t=2}^{T} \log \left(\frac{\boldsymbol{\alpha}_t \mathbf{1}_N^\top}{\boldsymbol{\alpha}_{t-1} \mathbf{1}_N^\top} \right) \\
&= \log \left(\frac{\boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{1}_N^\top}{\boldsymbol{\delta} \mathbf{1}_N^\top} \right) + \sum_{t=2}^{T} \log \left(\frac{\boldsymbol{\alpha}_{t-1}}{\boldsymbol{\alpha}_{t-1} \mathbf{1}_N^\top} \boldsymbol{\Gamma} \boldsymbol{P}(x_t) \mathbf{1}_N^\top \right) \\
&= \log \left(\boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{1}_N^\top \right) + \sum_{t=2}^{T} \log \left(\boldsymbol{\phi}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(x_t) \mathbf{1}_N^\top \right),
\end{aligned}$$

which is exactly stating, that the log-likelihood is a sum of logarithmic-values.

Furthermore, we implement working parameters to address constraints of positivity for σ and the t.p.m. row-sum constraint.
For the rest of this paragraph, denote by $\hat{\cdot}$ the estimator of some parameter $\cdot$.
Assume, for example, $N = 3$. Firstly, set the working parameters $\eta_i = \log \lambda_i$ for some parameter λ_i. After we have maximized the likelihood with respect to the unconstrained parameters, the constrained parameter estimates can be obtained by transforming back: $\hat{\lambda}_i = e^{\hat{\eta}_i}$. Next, start by defining the matrix with entries $\tau_{ij} \in \mathbb{R}$

$$\mathbf{T} = \begin{pmatrix} - & \tau_{12} & \tau_{13} \\ \tau_{21} & - & \tau_{23} \\ \tau_{31} & \tau_{32} & - \end{pmatrix},$$

and $g : \mathbb{R} \to \mathbb{R}^+$ (strictly increasing) function $g(x) = e^x$. Define

$$\nu_{ij} = \begin{cases} g(\tau_{ij}) & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases},$$

and

$$\gamma_{ij} = \frac{\nu_{ij}}{\sum_{k=1}^{N} \nu_{ik}}, \quad i, j = 1, 2, \ldots, N,$$

and $\mathbf{\Gamma} = (\gamma_{ij})_{i,j=1}^{N}$. We perform the calculation of the likelihood-maximizing parameters in two steps:

I. Maximize $\mathcal{L}_T$ with respect to the working parameters $\mathbf{T} = \{\tau_{ij}\}$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$ which are all unconstrained by construction.

II. Transform the estimates of the working parameters to estimates of the natural parameters:

$$\widehat{\mathbf{T}} \to \widehat{\mathbf{\Gamma}}, \quad \widehat{\boldsymbol{\eta}} \to \widehat{\boldsymbol{\lambda}}.$$

Consider $\mathbf{\Gamma}$ for the case $g(x) = e^x$ and $N \in \mathbb{N} \setminus \{1\}$. Here we have

$$\gamma_{ij} = \frac{\exp(\tau_{ij})}{1 + \sum_{k \neq i} \exp(\tau_{ik})}, \quad i \neq j,$$

and the diagonal elements of $\mathbf{\Gamma}$ follow from the row sums of 1. The transformation in the opposite direction is

$$\tau_{ij} = \log\left(\frac{\gamma_{ij}}{1 - \sum_{k \neq i} \gamma_{ik}}\right) = \log(\gamma_{ij}/\gamma_{ii}), \quad \text{i} \neq j.$$

2.2.5 Standard Errors & Confidence Intervals

Unfortunately, relatively little is known about the finite-sample properties of maximum-likelihood estimators in hidden Markov models (HMMs) [60, p. 56]. Nevertheless, under suitable regularity conditions, asymptotic theory is available. In practice, this typically requires an estimate of the variance-covariance matrix of the parameter estimators.

A common approach is to obtain standard errors from the (observed) Hessian of the log-likelihood evaluated at the maximiser. However, this procedure can become unreliable when one or more parameters lie on, or close to, the boundary of the admissible parameter space, a phenomenon that occurs frequently in applications [60, p. 56]. An alternative is parametric bootstrapping, but this is computationally demanding. Given the already substantial computational burden of the present thesis, we therefore restrict attention to Hessian-based inference.

Standard Errors via the Hessian Point estimates of the model parameters, $\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\lambda}})$, are straightforward to obtain by maximum likelihood. Exact (finite-sample) confidence intervals are, however, typically not available in closed form. Under suitable regularity conditions, the maximum-likelihood estimators (MLEs) in HMMs are consistent, asymptotically normal and asymptotically efficient [10, Chapter 12]. Consequently, approximate confidence intervals can be constructed from estimated standard errors via asymptotic normality.

It is nevertheless well documented that mixture-type models may require very large sample sizes for reliable inference and that components with small mixing weights, as well as overfitting through an excessive number of components, can lead to substantial practical difficulties [35, p. 68], [23, p. 53].

To estimate standard errors, we use the approximate Hessian of the negative log-likelihood at the optimum as returned by the `nlm` optimizer in `R`. Inverting this Hessian yields an estimate of the asymptotic variance-covariance matrix of the parameter estimators. Since the optimization is carried out in terms of transformed (unconstrained) working parameters $\boldsymbol{\eta}$, the reported Hessian corresponds to derivatives with respect to $\boldsymbol{\eta}$ rather than the original (natural) parameters $\boldsymbol{\zeta}$. Specifically, letting ℓ denote the log-likelihood, we obtain

$$\boldsymbol{H}_w = -\left(\frac{\partial^2 \ell}{\partial \eta_i \, \partial \eta_j}\right),$$

whereas we seek

$$\boldsymbol{H}_n = -\left(\frac{\partial^2 \ell}{\partial \zeta_i \, \partial \zeta_j}\right).$$

Following [40, p. 247], these matrices are related at the optimum by

$$\boldsymbol{H}_w = \boldsymbol{M}\,\boldsymbol{H}_n\,\boldsymbol{M}^\top \iff \boldsymbol{H}_n^{-1} = \boldsymbol{M}^\top\,\boldsymbol{H}_w^{-1}\,\boldsymbol{M}, \tag{21}$$

where $\boldsymbol{M}$ is the Jacobian of the transformation from working to natural parameters, with entries $m_{ij} = \partial\zeta_i/\partial\eta_j$. Since $\boldsymbol{M}$ is available in closed form for the transformations used, we obtain $\boldsymbol{H}_n^{-1}$ from $\boldsymbol{H}_w^{-1}$ via Equation 21 and then compute standard errors for the natural parameters. This procedure is applied provided the relevant parameters are not on, or extremely close to, the boundary of the admissible parameter space, in which case Hessian-based inference may be unreliable.

2.2.6 Forecasting, Decoding and State Prediction

In this section, $\boldsymbol{\delta}$ denotes the initial distribution, but every result is identical if it were to be the stationary distribution.

Conditional Densities Using the HMM likelihood formulation from Proposition 2.2, for $t = 2, 3, \ldots, T$ we obtain

$$\begin{aligned} f_{X_t|\mathbf{X}^{(-t)}}\left(x_t \mid \mathbf{x}^{(-t)}\right) &= \frac{\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{B}_2\cdots\boldsymbol{B}_{t-1}\boldsymbol{\Gamma}\boldsymbol{P}(x_t)\boldsymbol{B}_{t+1}\cdots\boldsymbol{B}_T\mathbf{1}_N^\top}{\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{B}_2\cdots\boldsymbol{B}_{t-1}\boldsymbol{\Gamma}\boldsymbol{B}_{t+1}\cdots\boldsymbol{B}_T\mathbf{1}_N^\top} \\ &\propto \boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma}\boldsymbol{P}(x_t)\boldsymbol{\beta}_t^\top, \end{aligned} \tag{22}$$

where $\boldsymbol{B}_t = \boldsymbol{\Gamma}\boldsymbol{P}(x_t)$, $\boldsymbol{\alpha}_t = \boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{B}_2\cdots\boldsymbol{B}_t$ and $\boldsymbol{\beta}_t^\top = \boldsymbol{B}_{t+1}\cdots\boldsymbol{B}_T\mathbf{1}_N^\top$.
For $t = 1$, we similarly have

$$\begin{aligned} f_{X_1|\mathbf{X}^{(-1)}}\left(x_1 \mid \mathbf{x}^{(-1)}\right) &= \frac{\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{B}_2\cdots\boldsymbol{B}_T\mathbf{1}_N^\top}{\boldsymbol{\delta}\boldsymbol{I}_{\boldsymbol{N}}\boldsymbol{B}_2\cdots\boldsymbol{B}_T\mathbf{1}_N^\top} \\ &\propto \boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{\beta}_1^\top. \end{aligned} \tag{23}$$

This ratio represents the conditional density of x_t, where the numerator corresponds to the joint likelihood with the observation at time t replaced by x_t, while the denominator is the full likelihood of the observed data, treating x_t as missing.

These conditional densities can be expressed as mixtures of the state-dependent densities. Since $\boldsymbol{P}(x) = \mathrm{diag}(f_1(x), \ldots, f_N(x))$, both Equation 22 and Equation 23 yield

$$f_{X_t|\mathbf{X}^{(-t)}}(x_t \mid \mathbf{x}^{(-t)}) \propto \sum_{i\in\mathcal{C}} d_i(t) f_{i,X_t}(x_t),$$

where for Equation 22, $d_i(t)$ equals the product of the i'th entry of $\boldsymbol{\beta}_t$ and the i'th entry of $\boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma}$, while for Equation 23, it is the product of the i'th entry of $\boldsymbol{\beta}_1$ and the i'th entry of $\boldsymbol{\delta}$. Normalizing

these weights gives

$$f_{X_t|\mathbf{X}^{(-t)}}\left(x_t \mid \mathbf{x}^{(-t)}\right) = \sum_{i\in\mathcal{C}} w_i(t) f_{i,X_t}(x_t), \qquad w_i(t) = \frac{d_i(t)}{\sum_{j\in\mathcal{C}} d_j(t)}.$$

Here, $w_i(t)$ are mixing weights that depend on the model parameters and on the remaining observations $\mathbf{x}^{(-t)}$.

Forecast Density Forecast densities are a special case of conditional densities. Let $h \in \mathbb{Z}^+$ denote the forecast horizon. For continuous-valued observations, the h-step-ahead forecast density $f_{X_{T+h}|\mathbf{X}^{(T)}}\left(x_{T+h} \mid \mathbf{x}^{(T)}\right)$ is obtained analogously to Equation 22:

$$\begin{aligned} f_{X_{T+h}|\mathbf{X}^{(T)}}\left(x_{T+h} \mid \mathbf{x}^{(T)}\right) &= \frac{f_{\mathbf{X}^{(T)},X_{T+h}}\left(\mathbf{x}^{(T)}, x_{T+h}\right)}{f_{\mathbf{X}^{(T)}}\left(\mathbf{x}^{(T)}\right)} \\ &= \frac{\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{B}_2\cdots\boldsymbol{B}_T\boldsymbol{\Gamma}^h\boldsymbol{P}(x_{T+h})\mathbf{1}_N^\top}{\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{B}_2\cdots\boldsymbol{B}_T\mathbf{1}_N^\top} \\ &= \frac{\boldsymbol{\alpha}_T\boldsymbol{\Gamma}^h\boldsymbol{P}(x_{T+h})\mathbf{1}_N^\top}{\boldsymbol{\alpha}_T\mathbf{1}_N^\top} \\ &= \boldsymbol{\phi}_T\boldsymbol{\Gamma}^h\boldsymbol{P}(x_{T+h})\mathbf{1}_N^\top, \qquad \boldsymbol{\phi}_T = \frac{\boldsymbol{\alpha}_T}{\boldsymbol{\alpha}_T\mathbf{1}_N^\top}. \end{aligned}$$

Thus, the forecast density is also a mixture of the N state-dependent densities:

$$f_{X_{T+h}|\mathbf{X}^{(T)}}\left(x_{T+h} \mid \mathbf{x}^{(T)}\right) = \sum_{i\in\mathcal{C}} \psi_i(h)\, f_{i,X_{T+h}}(x_{T+h}),$$

where $\psi_i(h)$ is the i'th entry of $\boldsymbol{\phi}_T\boldsymbol{\Gamma}^h$. Since the full forecast distribution is available, it is possible to construct both point and full interval forecasts. As the forecast horizon h increases, the predictive density converges to the marginal stationary density of the HMM, i.e.

$$\lim_{h\to\infty} f_{X_{T+h}|\mathbf{X}^{(T)}}(x_{T+h} \mid \mathbf{x}^{(T)}) = \lim_{h\to\infty} \boldsymbol{\phi}_T\boldsymbol{\Gamma}^h\boldsymbol{P}(x_{T+h})\mathbf{1}_N^\top = \boldsymbol{\delta}\boldsymbol{P}(x_{T+h})\mathbf{1}_N^\top, \tag{24}$$

where $\boldsymbol{\delta}$ is the stationary distribution of the Markov chain. The limit follows from the fact that for any nonnegative (row) vector $\boldsymbol{\vartheta}$ whose entries sum to 1, the vector $\boldsymbol{\vartheta}\boldsymbol{\Gamma}^h$ approaches $\boldsymbol{\delta}^*$ as $h \to \infty$, provided that the chain is irreducible and aperiodic [21, p. 394].

Decoding Proceeding, we infer the most likely sequence of Markov states that generated the observed data under the fitted model.
Firstly, rewrite the conditional distribution of C_t given the observations, for $i \in \mathcal{C}$ as

$$\begin{aligned}\mathbb{P}\left(C_t = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) &= \frac{f_{\mathbf{X}^{(T)},C_t}\left(\mathbf{x}^{(T)}, i\right)}{f_{\mathbf{X}^{(T)}}\left(\mathbf{x}^{(T)}\right)} \\ &= \frac{\alpha_t(i)\beta_t(i)}{\mathcal{L}_T}\end{aligned}$$

For each $t \in \{1, \ldots, T\}$, given the observations, the most probable state i_t^*, is defined as

$$i_t^* = \underset{i=1,\ldots,N}{\operatorname{argmax}} \mathbb{P}\left(C_t = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right). \tag{25}$$

In other words, this approach determines the most likely state, locally, separately for each time-t by maximizing the conditional probability. Hence the name "Local Decoding".

However, we are most interested in the most likely sequence globally of hidden states. Hence the name "Global Decoding". As such, we are interested in the quantity

$$(i_1^*, \ldots, i_T^*) = \underset{(i_1,\ldots,i_T)\in\mathcal{C}^T}{\operatorname{argmax}} \mathbb{P}\left(\mathbf{C}^{(T)} = \mathbf{c}^{(T)} \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right). \tag{26}$$

Finding the solution of Equation 26 over all possible state sequences involves N^T function evaluations which is not feasible. A feasible approach is the so called "Viterbi algorithm" [57, 22]. Define for $i \in \mathcal{C}$ and $t = 1$

$$\xi_{1i} = f_{X_1,C_1}(x_1, i) = \delta_i f_{i,X_1}(x_1),$$

and for $t = 2, 3, \ldots, T$,

$$\xi_{ti} = \max_{c_1,c_2,\ldots,c_{t-1}} f_{\mathbf{X}^{(t)},C_t,\mathbf{C}^{(t-1)}}\left(\mathbf{x}^{(t)}, i, \mathbf{c}^{(t-1)}\right).$$

This leads us to the recursion of ξ.

Proposition 2.5. *For $t = 2, 3, \ldots, T$ and $j \in \mathcal{C}$, it follows that*

$$\xi_{tj} = \left(\max_i(\xi_{t-1,i}\gamma_{ij})\right) f_{j,X_t}(x_t).$$

Proof. Fix $t \geq 2$ and $j \in \mathcal{C}$. For any $(c_1, \ldots, c_{t-1}) \in \mathcal{C}^{t-1}$, by the HMM conditional independences,

$$\begin{aligned} f_{\mathbf{X}^{(t)}, C_t, \mathbf{C}^{(t-1)}}(\mathbf{x}^{(t)}, j, \mathbf{c}^{(t-1)}) &= f_{j,X_t}(x_t)\mathbb{P}(C_t = j \mid C_{t-1} = c_{t-1}) \\ &\quad \times f_{\mathbf{X}^{(t-1)}, C_{t-1}, \mathbf{C}^{(t-2)}}(\mathbf{x}^{(t-1)}, c_{t-1}, \mathbf{c}^{(t-2)}) \\ &= f_{j,X_t}(x_t)\gamma_{c_{t-1}j} f_{\mathbf{X}^{(t-1)}, C_{t-1}, \mathbf{C}^{(t-2)}}(\mathbf{x}^{(t-1)}, c_{t-1}, \mathbf{c}^{(t-2)}). \end{aligned}$$

Maximizing over $\mathbf{c}^{(t-1)}$ and extracting the factors that do not depend on the maximization variables gives

$$\xi_{tj} = f_{j,X_t}(x_t) \max_{c_{t-1} \in \mathcal{C}} \left\{ \gamma_{c_{t-1}j} \max_{c_1, \ldots, c_{t-2} \in \mathcal{C}} f_{\mathbf{X}^{(t-1)}, C_{t-1}, \mathbf{C}^{(t-2)}} \left(\mathbf{x}^{(t-1)}, c_{t-1}, \mathbf{c}^{(t-2)} \right) \right\}.$$

By the definition of $\xi_{t-1,i}$,

$$\max_{c_1, \ldots, c_{t-2}} f_{\mathbf{X}^{(t-1)}, C_{t-1}, \mathbf{C}^{(t-2)}}(\mathbf{x}^{(t-1)}, i, \mathbf{c}^{(t-2)}) = \xi_{t-1,i},$$

so

$$\xi_{tj} = f_{j,X_t}(x_t) \max_{i \in \mathcal{C}} (\gamma_{ij} \xi_{t-1,i}),$$

which is the desired recursion. The initialization $\xi_{1i} = \delta_i f_{i,X_1}(x_1)$ follows from $f_{X_1,C_1}(x_1, i) = f_{i,X_1}(x_1)\mathbb{P}(C_1 = i)$. □

The required maximizing sequence of states $\{i\}_{i=1}^T$ can then be determined recursively from

$$i_T = \operatorname*{argmax}_{i=1,\ldots,N} \xi_{Ti},$$

and for $t = T-1, T-2, \ldots, 1$ from

$$i_t = \operatorname*{argmax}_{i=1,2,\ldots,N} (\xi_{ti} \gamma_{i,i_{t+1}}).$$

Because the global-decoding objective is a product of probabilities, it's convenient to maximize its logarithm to avoid numerical underflow; the Viterbi recursions translate directly to the log domain. As an alternative, we can use likelihood-style scaling by normalizing each time-t row of the matrix $\{\xi_{ti}\}$ so that the entries sum to 1. We adopt this approach. The Viterbi algorithm applies to both stationary and non-stationary (time-inhomogeneous) Markov chains, that is, the initial distribution need not be the stationary distribution.

State Prediction We turn our attention to conditional distributions of C_t when $t > T$, i.e. state prediction.

Proposition 2.6 (Filtering, smoothing and prediction)**.** *Let $\mathcal{C} = \{1, \ldots, N\}$. For each $t \in \{1, \ldots, T\}$ define the forward variables*

$$\alpha_t(i) := f_{\mathbf{X}^{(t)}, C_t}\left(\mathbf{x}^{(t)}, i\right), \qquad i \in \mathcal{C},$$

and the likelihood

$$\mathcal{L}_T := f_{\mathbf{X}^{(T)}}\left(\mathbf{x}^{(T)}\right) = \sum_{i \in \mathcal{C}} \alpha_T(i).$$

Let $\boldsymbol{\alpha}_T := (\alpha_T(1), \ldots, \alpha_T(N))$ be the $1 \times N$ row vector, let $\mathbf{\Gamma} = (\gamma_{ij})_{i,j=1}^N$ be the time-homogeneous transition matrix and let $\mathbf{e}_i \in \mathbb{R}^N$ denote the ith canonical basis column vector.

Then, for any $i \in \mathcal{C}$,

$$\mathcal{L}_T \, \mathbb{P}\left(C_t = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \begin{cases} \boldsymbol{\alpha}_T \, \mathbf{\Gamma}^{t-T} \, \mathbf{e}_i, & t > T \quad \textit{state prediction,} \\ \alpha_T(i), & t = T \quad \textit{filtering,} \\ \alpha_t(i) \; \beta_t(i), & 1 \leq t < T \quad \textit{smoothing,} \end{cases}$$

where $\beta_t(i)$ are the backward variables from the forward-backward algorithm, with $\beta_T(i) = 1$ for all $i \in \mathcal{C}$.

Equivalently, letting $\boldsymbol{\phi}_T := \boldsymbol{\alpha}_T / \mathcal{L}_T$ denote the filtered state distribution at time T, for $h \geq 0$,

$$\mathbb{P}\left(C_{T+h} = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \boldsymbol{\phi}_T \, \mathbf{\Gamma}^h \, \mathbf{e}_i.$$

Proof. Fix $h \geq 1$ and $i \in \mathcal{C}$. By the law of total probability over C_T,

$$\mathbb{P}\left(C_{T+h} = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \sum_{i \in \mathcal{C}} \mathbb{P}\left(C_{T+h} = i \mid C_T = j, \; \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) \mathbb{P}\left(C_T = j \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right).$$

By the Markov property of $\{C_t\}$ and the HMM conditional-independence structure, the future of the chain is independent of the past observations given C_T, hence

$$\mathbb{P}\left(C_{T+h} = i \mid C_T = j, \; \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \mathbb{P}\left(C_{T+h} = i \mid C_T = j\right) = \left(\mathbf{\Gamma}^h\right)_{ji}.$$

Therefore,

$$\mathbb{P}\left(C_{T+h} = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \sum_{i \in \mathcal{C}} \left(\mathbf{\Gamma}^h\right)_{ji} \mathbb{P}\left(C_T = j \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right).$$

Moreover,

$$\mathbb{P}\left(C_T = j \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \frac{f_{\mathbf{X}^{(T)}, C_T}\left(\mathbf{x}^{(T)}, j\right)}{f_{\mathbf{X}^{(T)}}\left(\mathbf{x}^{(T)}\right)} = \frac{\alpha_T(j)}{\mathcal{L}_T}.$$

Letting $\boldsymbol{\phi}_T := (\phi_T(1), \ldots, \phi_T(N))$ with $\phi_T(j) = \alpha_T(j)/\mathcal{L}_T$, we obtain

$$\mathbb{P}\left(C_{T+h} = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \boldsymbol{\phi}_T \boldsymbol{\Gamma}^h \mathbf{e}_i.$$

Multiplying by $\mathcal{L}_T$ yields the equivalent form

$$\mathcal{L}_T \mathbb{P}\left(C_{T+h} = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \boldsymbol{\alpha}_T \boldsymbol{\Gamma}^h \mathbf{e}_i,$$

which is the state-prediction case.

For $h = 0$, $\boldsymbol{\Gamma}^0 = \mathbf{I}$, so

$$\mathbb{P}\left(C_T = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \phi_T(i) = \frac{\alpha_T(i)}{\mathcal{L}_T},$$

equivalently $\mathcal{L}_T \mathbb{P}\left(C_T = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \alpha_T(i)$, which is the filtering identity.

For $1 \leq t < T$, the standard forward-backward identity gives

$$\mathbb{P}\left(C_t = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) = \frac{\alpha_t(i)\,\beta_t(i)}{\mathcal{L}_T},$$

and since $\beta_T(i) = 1$ by initialization of the backward recursion, the smoothing expression matches the filtering expression at $t = T$. This completes the proof. □

Note that as $h \to \infty$, $\boldsymbol{\phi}_T \boldsymbol{\Gamma}^h \to \boldsymbol{\delta}$ and so $\mathbb{P}\left(C_{T+h} = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right) \to \delta_i$.

2.2.7 Number of States

HMMs are prone to overfitting [26, p. 2]. That is, HMMs are not well suited for order estimation as small variations in the data are known to cause such models to overestimate the number of groups as well as the frequency of transitions when the number of states is unknown. This raises the question; How does one adequately choose the number of states in a HMM?

In HMMs, the number of states must be specified a priori to the analysis rather than estimated during model fitting by the above-mentioned reason. However, this decision can be challenging, as standard model selection criteria like AIC and BIC often favour a large number of states, which can reduce interpretability[2]. In particular, AIC tends to select more states as increased model flexibility allows for better data fitting, though this can come at the expense of generalizability and interpretability. In other words, we might misspecify some variation in the data as an extra false state-i, as it gives a better model fitting, even though it might just be a (large) variation

[2]This will become evident in Section 2.4.1.

within a true state-j. [43] and [39, p. 4] highlighted this issue, recommending that the choice of states should be guided by domain expertise and model validation rather than relying solely on selection criteria. As such, our information criteria, AIC and BIC, will also be utilized for model selection, but not exclusively. We describe the information criteria in Section 2.4.1.

A proposed solution to the a priori number of state selection, is the heuristic method of counting modes in the distribution of the data. However, this can be severely problematic. For example, assume the known number of states is two. If the means are approximately equal but the variance differ it is virtually impossible by visual examination to determine that the number of states is one or some larger integer.

Following [43] and [39], we determine the number of states based on domain expertise only. However, we note that the application of HMMs to financial contexts, and especially to interest rates, remains extremely limited. Consequently, we must develop our own arguments to justify the chosen number of states based on domain expertise.

In the context of equity market modeling, selecting between 1 and 5 states provides a meaningful balance between model complexity, interpretability and macro-financial relevance. A two-state model may capture broad bull and bear market regimes. A third state can correspond to recovery or neutral phases in market cycles. Higher state counts may reflect nuanced market phases, such as asset bubbles, mild corrections, or crashes.

The number of states affects the parameter estimation in a regime-switching Black-Scholes model. In particular, the volatility parameter σ becomes state-dependent. If too many states are included, temporary fluctuations in volatility may be mistaken for persistent regime shifts. Conversely, too few states may obscure significant differences in market regimes, such as between stable bull markets and volatile rallies.

We interpret (or rather hypothesize) the possible values of $N \in \{1, 2, 3, 4, 5\}$ as follows:

- $N = 1$: The trivial case, meaning, no regime-switching. The standard Black-Scholes model with constant drift and volatility.
- $N = 2$: A dichotomy of bull and bear markets, capturing broad upturn and downturn market regimes.
- $N = 3$: Extension to classical business cycle phases: expansion, recession and recovery.
- $N = 4$: Potential to differentiate mild versus severe market states, e.g., modest bull markets vs. overheated bubbles, or shallow vs. deep downturns. a 4-state system could also be nuances of a bull and bear market, i.e. two bull and two bear market states that allow for varying degrees of severity.
- $N = 5$: Allows capturing even finer distinctions, such as neutral/stagnant markets or extreme panic phases during financial crises.

For example, during a moderate downturn, the drift μ may decline and volatility σ rise modestly, reflecting controlled risk aversion. In contrast, in an extreme crisis, μ becomes strongly negative and σ spikes due to panic-driven trading, credit contractions and liquidity crises. Capturing both with the same state may understate risk in crisis scenarios or overstate it in moderate corrections.

In summary, based on economic reasoning and business cycle theory, we restrict our analysis to $N \in \{1, 2, 3, 4, 5\}$. This reflects plausible macroeconomic regimes and aims to avoid overfitting while maintaining explanatory power and interpretability in financial modeling.

2.2.8 Simulation

We simulate the BS-HMM with parameters $n = 25000$ and daily observations $\Delta = 1/252$ (approximately $25000/252 \approx 99$ years). The model parameters are seen in Table 2. To simulate from a N-state hidden Markov model, we extend the Euler discretized version of the BS SDE given in Equation 11 to include a state sequence from a simulated Markov chain, simply by using the `sample()`-function in base `R`. Combining the simulated Markov chain with the discretized BS SDE, we achieve the state-dependent Euler-discretized version of the BS SDE

$$\hat{S}_{t+\Delta} = \hat{S}_t + \mu_i \hat{S}_t \Delta + \sigma_i \hat{S}_t \sqrt{\Delta} Z^{\mathbb{P}}, \quad i \in \mathcal{C}.$$

We limit ourself to the most computationally dragging case, which is the 5-state BS-HMM. The results are seen in Table 2. The simulated price path is shown in Figure 11. `nlm` in `R` is used to maximize the likelihood given in Equation 20 for the 5-state BS-HMM with both μ and σ assumed state-dependent.

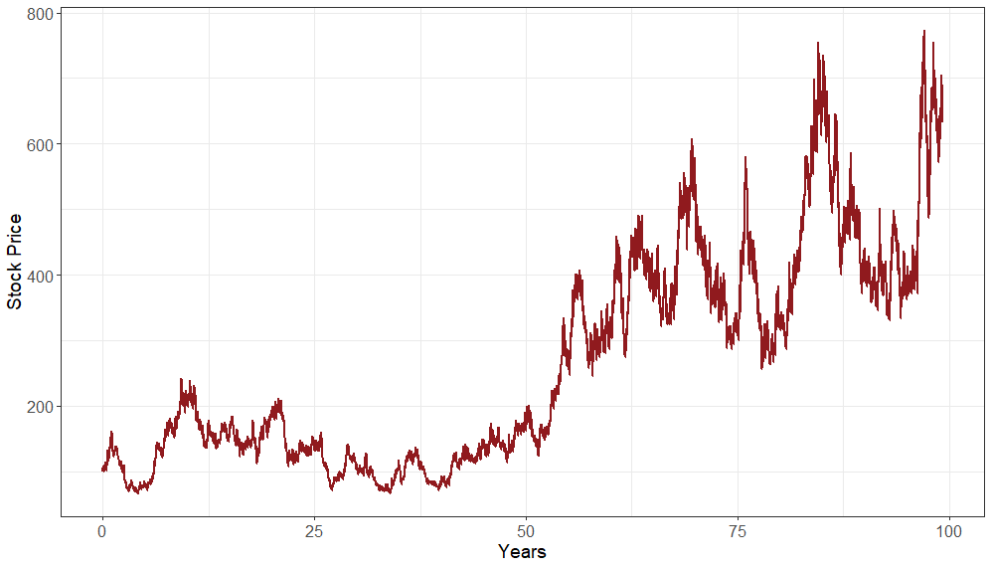


Figure 11: *BS-HMM simulated stock price path.*

Parameter	True Values	Estimated Values	Relative Error (%)
$\boldsymbol{\mu}$	$\begin{bmatrix} -0.2000 \\ -0.06000 \\ 0.03000 \\ 0.08000 \\ 0.1200 \end{bmatrix}$	$\begin{bmatrix} -0.07963 \\ 0.02766 \\ 0.06999 \\ -0.03065 \\ 0.07887 \end{bmatrix}$	$\begin{bmatrix} 60.19 \\ 146.1 \\ 133.3 \\ 138.3 \\ 34.27 \end{bmatrix}$
$\boldsymbol{\sigma}$	$\begin{bmatrix} 0.3000 \\ 0.2400 \\ 0.1800 \\ 0.1500 \\ 0.2200 \end{bmatrix}$	$\begin{bmatrix} 0.3046 \\ 0.1779 \\ 0.1655 \\ 0.1104 \\ 0.2409 \end{bmatrix}$	$\begin{bmatrix} 1.530 \\ 25.89 \\ 8.043 \\ 26.38 \\ 9.520 \end{bmatrix}$
$\boldsymbol{\delta}$	$\begin{bmatrix} 0.09973 \\ 0.1552 \\ 0.2049 \\ 0.2498 \\ 0.2904 \end{bmatrix}$	$\begin{bmatrix} 0.08005 \\ 0.3991 \\ 0.1334 \\ 0.03329 \\ 0.3542 \end{bmatrix}$	$\begin{bmatrix} 19.73 \\ 157.2 \\ 34.90 \\ 86.67 \\ 21.95 \end{bmatrix}$
$\boldsymbol{\Gamma}$	$\begin{bmatrix} 0.8895 & 0.01639 & 0.02388 & 0.03137 & 0.03886 \\ 0.008912 & 0.8912 & 0.02580 & 0.03330 & 0.04081 \\ 0.01081 & 0.01833 & 0.8929 & 0.03524 & 0.04276 \\ 0.01271 & 0.02024 & 0.02778 & 0.8945 & 0.04473 \\ 0.01462 & 0.02217 & 0.02972 & 0.03726 & 0.8962 \end{bmatrix}$	$\begin{bmatrix} 0.9058 & 0.09336 & 0.0004770 & 1.377 \times 10^{-8} & 0.0003406 \\ 0.01883 & 0.9024 & 0.001744 & 0.0002923 & 0.07671 \\ 7.312 \times 10^{-7} & 0.00007296 & 0.9529 & 0.04707 & 3.392 \times 10^{-6} \\ 0.0006466 & 0.0001163 & 0.0001677 & 0.7755 & 0.2236 \\ 0.00008703 & 0.08942 & 0.01665 & 2.394 \times 10^{-6} & 0.8938 \end{bmatrix}$	$\begin{bmatrix} 1.833 & 469.8 & 98.00 & 100.0 & 99.12 \\ 111.2 & 1.262 & 93.24 & 99.12 & 87.99 \\ 99.99 & 99.60 & 6.720 & 33.55 & 99.99 \\ 94.91 & 99.43 & 99.40 & 13.31 & 399.8 \\ 99.40 & 303.4 & 43.97 & 99.99 & 0.2666 \end{bmatrix}$

Table 2: *True, estimated and relative error (%) for the Black-Scholes 5-state hidden Markov model parameters where μ and σ are modeled as state-dependent.*

In the 5-state BS-HMM with both state-dependent drift and volatility, occasional instability of the MLEs in simulation can be traced to an interaction between Euler discretizations error and regime switching. Under switching, the coefficients $(\mu_{C_t}, \sigma_{C_t})$ change discontinuously whenever the hidden chain transitions from state i to state j. This interrupts the usual accumulation of discretizations error. In other words, each regime segment behaves like a new local diffusion with different parameters and frequent switching effectively prevents the Euler scheme from settling into its small-step asymptotics. In this sense, state transitions repeatedly reset the local approximation problem, which can throttle convergence.

Discretizations error is present for all models, but it is most pronounced in the BS-HMM because the likelihood is highly sensitive to small regime-dependent perturbations in the simulated return distribution. Consequently, modest path-level simulation errors can be amplified into materially different inferred regimes and, in turn, into numerically problematic MLEs. A rigorous convergence analysis for discretized Markov-switching diffusions is outside the scope of this thesis.

2.3 Continuous State-Space Models

The structure of states in HMMs can be problematic. As explored in Section 2.2.7, the number of states is rarely known a priori. As such, one has to rely on domain expertise to specify the number of states. Model selection criteria and examination of pseudo-residuals is the preferred method. However, in fact, most of the time the number of states remains difficult in practice when the underlying data generating process is unknown. The method has a major drawback as we would not know wether to fit models from $N = 2$ to $N = 5$ or $N = 100$ Furthermore, as the number of states can possibly be extremely large, the number of parameters rise extremely fast. If we for example have 10 states BS-HMM and 2 state-dependent variables, we would need to estimate a

total of $10^2 + 10 = 110$ parameters for the 10-state HMM with 2 state-dependent parameters. As such, it can be advantageous to consider alternative model formulations, where the state process is continuous-valued as opposed to discrete-valued. Furthermore, such continuous state space models are simpler in terms of parameter count.

2.3.1 Autoregressive Processes

For the S&P 500 data, it could be intuitive to assume that the rate of occurrences is continuous-valued. A continuous state-space would allow for gradual, or rather, less abrupt changes over the years. Furthermore, such model would be able to adapt to changes in Algo-trading, economical policies and human behavior, which is a major drawback of the BS-HMM. A simple model that would capture such changes could be formulated as:
(BS-SSM):

$$\begin{aligned} C_t &= \rho\, C_{t-1} + \varepsilon_t, && \varepsilon_t \sim \mathcal{N}\big(0, \sigma_\varepsilon^2\big), \\ \mu_t &= \mu\, e^{C_t}, \\ \sigma_t &= \sigma\, e^{C_t}, \\ X_t \mid C_t &\sim \mathcal{N}\left(\left(\mu_t - \tfrac{1}{2}\sigma_t^2\right)\,\Delta,\ \sigma_t^2\Delta\right), \end{aligned} \tag{27}$$

or by factor loading the latent states by constants $\beta_\mu, \beta_\sigma \in \mathbb{R}$:
(BS-SSM$_\beta$):

$$\begin{aligned} C_t &= \rho\, C_{t-1} + \varepsilon_t, && \varepsilon_t \sim \mathcal{N}\big(0, \sigma_\varepsilon^2\big), \\ \mu_t &= \mu + \beta_\mu\, C_t, \\ \sigma_t &= \sigma\, \exp\big(\beta_\sigma C_t\big), \\ X_t \mid C_t &\sim \mathcal{N}\left(\left(\mu_t - \tfrac{1}{2}\sigma_t^2\right)\Delta,\ \sigma_t^2\Delta\right), \end{aligned} \tag{28}$$

for $t = 1, 2, ...$ and with the recursion initiated in $C_0 = C \in \mathbb{R}_+$[3]. The autoregressive parameter is $\rho \in \mathbb{R}$ and the innovations ε_t are independently and identically distributed with a normal distribution with mean zero and variance $\sigma_\varepsilon^2 > 0$. The subscript ε in σ_ε to avoid confusion with the BSM parameter σ. In other words, ε_t are i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$. This is called an autoregressive process of order 1. It follows that $\mathbb{E}[C_t \mid C_{t-1}] = \rho C_{t-1}$ while $\mathbb{V}[C_t \mid C_{t-1}] = \sigma_\epsilon^2$. As such, the dynamics are modeled through the conditional mean of C_t given the past.

Consider the simple recursion for C_t

$$C_t = \rho^t C + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i}.$$

[3]Note that we index from 0 for the AR(1) theory rather than 1 as the HMM theory. This is to adhere to the general literature. It is a simple shift of index.

In particular, by $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, C_t is normally distributed with time-varying parameters

$$\begin{aligned}\mathbb{E}[C_t] &= \rho^t C \\ \mathbb{V}[C_t] &= \left(1 + \rho^2 + \rho^4 + ... + \rho^{2(t-1)}\right) \sigma_\varepsilon^2\end{aligned}$$

If $|\rho| < 1$, $\mathbb{E}[C_t] \to 0$ as $\rho^t \to 0$ for $t \to \infty$. Concludingly, for $|\rho| < 1$, as $t \to \infty$, C_t will resemble the so called linear-process,

$$C_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i},$$

in terms of the sequence $\{\varepsilon_t\}_{t=...,-1,0,1,...}$ of i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$ variables. The process $\{C_t^*\}_{t=0,1,...}$ is then Gaussian distributed with

$$\begin{aligned}\mathbb{E}[C_t^*] &= 0, \\ \mathbb{V}[C_t^*] &\overset{\dagger}{=} \frac{\sigma_\varepsilon^2}{(1-\rho^2)},\end{aligned}$$

as

$$\frac{1-\rho^{2t}}{1-\rho^2}\sigma_\varepsilon^2 \overset{\dagger\dagger}{\to} \frac{\sigma_\varepsilon^2}{(1-\rho^2)}, \quad t \to \infty,$$

where † and †† follows from Lemma A.2.6. The distribution of C_t^* is evidently independent of time and is an example of a stationary process. Thus, if $|\rho| < 1$, C_t is said to be asymptotically stationary in the sense that it resembles the stationary process C_t^* for $t \to \infty$[4].

Stationarity & Distributions Proceeding, we formalize the notion of stationarity and examine distributional properties of the AR(1) process by introducing the notion of a drift function.

Definition 2.4. *The process $\{C_t\}_{t=0,1,...}$ is said to be a stationary process if for all $t, h \geq 0$, the joint distribution of $(C_t, \ldots, C_{t+h})$ does not depend on $t \geq 0$.*

By Definition 2.4, note that for a stationary process with well-defined second order moments, $\mathbb{E}[C_t]$ and $\mathbb{V}[C_t]$ are constant and that the covariance between C_t, C_{t+h}, i.e. $\text{Cov}[C_t, C_{t+h}]$ depends only on h and not t.

The definition of stationarity comments only on the joint distribution of the variables. It states nothing about dependence over time. Assume $\{C_t\}_{t\in\mathbb{Z}}$ is a stationary process that is dependent over time with finite second order moment $\mathbb{E}\left[|C_t|^2\right] < \infty$. A often used indicator to detect dependence is the auto-correlation. For a stationary process $C_t \in \mathbb{R}$, the auto-covariance function

[4]We will define the concept of stationarity in the next paragraph.

is given by

$$v(h) = \text{Cov}[C_t, C_{t+h}],$$

and ACF defined by,

$$\text{ACF}(h) = \text{Corr}[C_t, C_{t+h}] = \frac{\text{Cov}[C_t, C_{t+h}]}{\sqrt{\mathbb{V}[C_t]\mathbb{V}[C_{t+h}]}} \overset{\dagger}{=} \frac{v(h)}{v(0)},$$

where † holds by stationarity. The function for various h describe the correlation and hence indicate dependence over time.

We now define weak dependence and a result for time series that are (assumed) weakly dependent.

Definition 2.5. *Let $\{C_t\}_{t\in\mathbb{Z}}$ be a stationary process with $\mathbb{E}\left[|C_t|^2\right] < \infty$ and auto-covariance function*

$$v(h) = \text{Cov}(C_t, C_{t+h}), \quad h \in \mathbb{Z}.$$

The process is said to be weakly dependent (or short-range dependent) if

$$\sum_{h=-\infty}^{\infty} |v(h)| < \infty.$$

Equivalently, since $v(-h) = v(h)$ for a stationary process,

$$\sum_{h=0}^{\infty} |v(h)| < \infty.$$

Lemma 2.1. *Let $\{C_t\}_{t\in\mathbb{Z}}$ be stationary with $\mathbb{E}[C_t] = \mu$ and $\mathbb{E}\left[|C_t|^2\right] < \infty$. Let $v(h) = \text{Cov}(C_t, C_{t+h})$ denote the auto-covariance function. Define the sample mean $\bar{C}_T = \frac{1}{T}\sum_{t=1}^{T} C_t$ and suppose that*

$$\sum_{h=-\infty}^{\infty} |v(h)| < \infty.$$

Then the limit

$$\Omega := v(0) + 2\sum_{h=1}^{\infty} v(h)$$

exists and is finite and moreover

$$\mathbb{V}\left(\sqrt{T}\left(\bar{C}_T - \mu\right)\right) \longrightarrow \Omega, \qquad T \to \infty.$$

Proof. Set $Y_t = C_t - \mu$, so that $\mathbb{E}[Y_t] = 0$ and $\text{Cov}(Y_t, Y_{t+h}) = v(h)$ by stationarity. Let $Z_T =$

$\sum_{t=1}^{T} Y_t$. Then

$$\mathbb{V}(Z_T) = \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}(Y_t, Y_s) = \sum_{t=1}^{T}\sum_{s=1}^{T} v(s-t).$$

Now group the double sum by lags $h = s - t$. The diagonal terms contribute $Tv(0)$. For each $h \in \{1, \ldots, T-1\}$ there are exactly $T - h$ pairs (t, s) with $s - t = h$, hence

$$\mathbb{V}(Z_T) = Tv(0) + 2\sum_{h=1}^{T-1} (T-h)\, v(h).$$

Therefore,

$$\begin{aligned}\mathbb{V}\left(\sqrt{T}\left(\bar{C}_T - \mu\right)\right) = \mathbb{V}\left(\frac{1}{\sqrt{T}} Z_T\right) &= \frac{1}{T}\mathbb{V}(Z_T) \\ &= v(0) + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right) v(h).\end{aligned}$$

Since $\left|1 - \frac{h}{T}\right| \leq 1$ and $\sum_{h=1}^{\infty} |v(h)| < \infty$, we can pass to the limit by a standard tail-splitting argument. Fix $H \in \mathbb{N}$ and write

$$\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right) v(h) = \sum_{h=1}^{H}\left(1 - \frac{h}{T}\right) v(h) + \sum_{h=H+1}^{T-1}\left(1 - \frac{h}{T}\right) v(h).$$

For fixed H, the first sum converges to $\sum_{h=1}^{H} v(h)$ as $T \to \infty$ and the second sum is bounded in absolute value by $\sum_{h=H+1}^{\infty} |v(h)|$, which can be made arbitrarily small by taking H large. Hence,

$$\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right) v(h) \;\longrightarrow\; \sum_{h=1}^{\infty} v(h),$$

and the claim follows. □

Accordingly, "mixing" (i.e., asymptotic independence) captures that the dependence between C_t, C_{t+h} vanishes as $h \to \infty$. This idea is crucial for time series and replaces the concept of independence. The idea is that a stationary process $\{C_t\}_{t=0,1,\ldots}$ is said to be mixing[5] (or, ergodic) if for all t, h and sets A, B,

$$\mathbb{P}((C_0, \ldots C_t) \in A, (C_h, \ldots, C_{t+h}) \in B) \to \mathbb{P}((C_0, \ldots, C_t) \in A)\mathbb{P}((C_0, \ldots, C_t) \in B), \quad h \to \infty$$

The notion is intuitively, that events removed far in time from one another are independent.

[5]The literature differs a lot on the notion on mixing; some even include different kinds of mixing (α, β-mixing.

Importantly, they imply that various Laws of Large Numbers apply.

We now turn our attention to the drift criterion. The criterion establishes conditions under which LLNs and Central Limit Theorems (CLTs) hold for time series. Let $\{C_t\}_{t=0,1,\dots}$ be a Markov chain that satisfies the drift criterion. The first implication of satisfying said criterion is that the initial value, C_0, can be assigned a distribution such that C_t is stationary. The second implication is finiteness of certain moments for the stationary version. Moreover, variations of LLN and CLT can be applied. Let $\{C_t\}_{t=0,1,\dots}$ denote a AR(1) process. Then, the distribution of $C_t \mid (C_{t-1}, \dots, C_0)$, $t \geq 1$ depends only on C_{t-1}, meaning, $C_t \mid C_{t-1} \sim \mathcal{N}(C_{t-1}\rho, \sigma_\varepsilon^2)$. As can be seen, the conditional distribution is Gaussian, which has some attractive properties.

We now state 2 assumptions based on [53, 38].

Assumption 2.6. *Assume that for $\{C_t\}_{t=1,0,\dots}$ with $C_t \in \mathbb{R}^p$ it holds that:*

(i) The conditional distribution of C_t given $(C_{t-1}, C_{t-2}, \dots, C_0)$ depends only on C_{t-1}, that is

$$C_t \mid C_{t-1}, C_{t-2}, \dots, C_0 \overset{d}{=} C_t \mid C_{t-1}.$$

(ii) The conditional distribution of C_t given C_{t-n}, for some $n \geq 1$, has a positive (n-step) conditional density $f(y \mid x) > 0$, which is continuous in both arguments.

(i) in Assumption 2.6 implies that $\{C_t\}_{t=0,1,\dots,}$ is a Markov chain on $\mathbb{R}^p$, or sometimes called a Markov chain on a general state space.

Example: For the purpose of our analysis, consider the AR(1) process. As ε_t are i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$ and independent of $(C_{t-1}, \dots, C_0)$, C_t conditional on $(C_{t-1}, \dots, C_0)$ has density

$$f_{C_t|C_{t-1}}(c_t \mid c_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{(c_t - \rho c_{t-1})^2}{2\sigma_\varepsilon^2}\right),$$

which depends exclusively on C_{t-1} and is well known to be positive and continuous in both arguments.

Next, define a drift function that satisfies Assumption 2.6. A drift function for some time series $\{C_t\}$, is some function $\delta(C_t) \geq 1$ which is not identically ∞. The role of a drift function is to measure the dynamical drift of C_t by studying the dynamics of the corresponding drift function $\delta(C_t)$. That is, we are interested in $\mathbb{E}[\delta(C_t) \mid C_{t-m}]$ for some $m \geq 1$.

Example: Consider again the AR(1) process with drift function $\delta(C_t) = 1 + C_t^2$ and $m = 1$.

Then, using that C_{t-1} and ε_t are independent, we obtain

$$\begin{aligned}\mathbb{E}[\delta(C_t) \mid C_{t-1}] &= \mathbb{E}[1 + (\rho C_{t-1} + \varepsilon_t)^2 \mid C_{t-1}] \\ &= 1 + \rho^2 \mathbb{E}[C_{t-1}^2 \mid C_{t-1}] + 2\rho C_{t-1}\mathbb{E}[\varepsilon_t \mid C_{t-1}] + \mathbb{E}\left[\varepsilon_t^2 \mid C_{t-1}\right] \\ &= 1 + \rho^2 C_{t-1}^2 + 2\rho C_{t-1}\mathbb{E}[\varepsilon_t] + \mathbb{E}\left[\varepsilon_t^2\right] \\ &= 1 + \sigma^2 + \rho^2 C_{t-1}^2 \\ &= \rho^2 \delta(C_{t-1}) + c, \quad c = (1 - \rho^2 + \sigma_\varepsilon^2).\end{aligned}$$

Thus we obtain a process that mimics a AR(1) process in $\delta(C_t)$, apart from some constant c. In other words,

$$\delta(C_t) = \rho^2 \delta(C_{t-1}) + c + \eta_t,$$

with $\eta_t = (\delta(C_t) - \mathbb{E}[\delta(C_t) \mid C_{t-1}])$ such that $\mathbb{E}[\eta_t] = 0$. As such, if $\rho^2 < 1$, $\delta(C_t)$ resembles a stationary AR(1) process. This leads us to the final assumption.

Assumption 2.7. *Assume that $\{C_t\}_{t=0,1,\dots}$, with $C_t \in \mathbb{R}^p$, satisfies Assumption 2.6. With drift function δ, $\delta(C_t) \geq 1$, assume that there exist positive constants M, C and φ with $\varphi < 1$, such that for some $m \geq 1$,*

(i) $\mathbb{E}\left[\delta(C_{t+m}) \,\middle|\, C_t = C\right] \leq \varphi\delta(C), \quad$ *for* $\|C\| > M$,

(ii) $\mathbb{E}\left[\delta(C_{t+m}) \,\middle|\, C_t = C\right] \leq C < \infty, \quad$ *for* $\|C\| \leq M$.

Example: Consider again the AR(1). To obtain the desired properties for C_t by making restrictions on ρ, we apply the drift criterion with $\delta(C_t) = 1 + C_t^2$. From the previous example

$$\mathbb{E}[\delta(C_t) \mid C_{t-1} = C] = \rho^2 \delta(C) + c, \quad c = (1 - \rho^2 + \sigma_\varepsilon^2).$$

Since

$$\lim_{|C|\to\infty} \frac{\mathbb{E}[\delta(C_t) \mid C_{t-1} = C]}{\delta(C)} = \rho^2,$$

we must require that $|\rho| < 1$ for the existence of positive constants M, φ with $\varphi < 1$, such that $\mathbb{E}[\delta(C_t) \mid C_{t-1} = C] \leq \varphi\delta(C)$ for $|C| > M$. By continuity of $\delta(C)$ and c, it automatically follows that $\mathbb{E}[\delta(C_t) \mid C_{t-1} = C] \leq C$ for some $C > 0$ for $|y| \leq M$. In other words, if $|\rho| < 1$, the AR(1) process satisfies the drift criterion in Assumption 2.7.

We are now in a position to list the main result of the section from [44, Thm. I.4.1], inspired by [27, 53].

Theorem 2.8. *Assume that $\{C_t\}_{t\geq 1}$ satisfies Assumption 2.7 with drift function δ. Then C_0 can be given an initial distribution such that C_t initiated in C_1 is stationary. With C_t denoting the*

stationary version, we have $\mathbb{E}\big[\delta(C_t)\big] < \infty$. *Moreover,* C_t *is mixing in the sense that, for any initial value* C_0, *the LLN Lemma 2.2 [44, p. 17] applies.*

Lemma 2.2. *Assume that with* $C_t \in \mathbb{R}^p$, $\{C_t\}_{t=0}^T$ *is a geometrically ergodic Markov chain with stationary solution* $\{C_t^*\}$. *Assume furthermore that the function* $g : \mathbb{R}^{p(m+1)} \to \mathbb{R}$, $m \geq 0$, *satisfies* $\mathbb{E}\left[g(C_t^*, C_{t-1}^*, \ldots, C_{t-m}^*)\right] < \infty$, *then as* $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} g(C_t, C_{t-1}, \ldots, X_{t-m}) \xrightarrow{\mathcal{P}} \mathbb{E}\left[g(C_t^*, C_{t-1}^*, \ldots, C_{t-m}^*)\right].$$

Unit Roots The underlying assumption of the before-mentioned analyses relied on the fact that $|\rho| < 1$. As stated in [44, Part II, p. 3], often met in the analysis of stock prices, it will be the case that $\rho = 1$. However, we shall shortly argue on why $\rho = 1$, i.e. the unit root case and cointegration, is not examined further in this thesis. When $\rho = 1$, it follows that the AR(1) process with $C_0 = C$ fixed, ε_t i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$,

$$C_t = C_{t-1} + \varepsilon_t = \sum_{i=1}^{t} \varepsilon_i + C. \tag{29}$$

In other words, C_t is the sum of a random walk $\sum_{i=1}^t \varepsilon_i$ and the initial value C. When $\rho = 1$, x_t is not stationary, not even asymptotically. Indeed, ε_t i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$, leading to the fact that the variance of C_t in Equation 29 is given by

$$\mathbb{V}\left[C_t\right] = \mathbb{V}\left[\sum_{i=1}^{t} \varepsilon_i + C\right] = t\sigma_\varepsilon^2,$$

which is increasing in t. Furthermore, note that

$$\Delta C_t = C_t - C_{t-1} = \sum_{i=1}^{t} \varepsilon_i + C - \sum_{i=1}^{t-1} \varepsilon_i + C = \varepsilon_t,$$

implying that the differenced process is stationary. Unit root analysis provides a framework to discriminate between the pair. The former is a non-stationary random walk case (the hypothesis of non-stationarity) and the latter a stationary case (the hypothesis of stationarity).

It is now apparent why the raw S&P 500 data was transformed to returns. However, note that we assumed that the state process is modeled through an autoregressive process of order 1 and not the asset prices.

A unit-root state behaves like a random walk, drifting without pullback. This destroys a stable baseline for "high/low" regimes and undermines interpretability. Furthermore, the state's uncertainty grows with sample length, such that there is no steady-state signal-to-noise. If the

state feeds the mean or (especially) the variance, implied moments can blow up over time. For stable inference and meaningful regimes, we therefore impose $|\rho| < 1$. Furthermore, as the state-process is unobserved, we have no means of testing for the hypothesis of non-stationarity via a Dickey-Fuller test [14] for the state-process.

2.3.2 Likelihood Formulation & Parameter Estimation

We consider the basic SSM in which the state process is univartiate. Such a SSM is characterized by two prcoesses (almost identical to that of the discrete-valued hidden Markov model)[6]:

1. A continuous-valued hidden Markov state process, $\{C_t\}_{t\in\mathbb{N}}$.
2. A observed process, $\{X_t\}_{t\in\mathbb{N}}$, whose realizations are assumed to be conditionally independent, given the states.

Formally, for a density function f, the assumptions can be formalized as:

$$f_{C_t|\mathbf{C}^{(t-1)}}\left(c_t \mid \mathbf{c}^{(t-1)}\right) = f(c_t \mid c_{t-1}), \quad t = 2, 3, \ldots,$$
$$f_{X_t|\mathbf{X}^{(t-1)},\mathbf{C}^{(t)}}\left(x_t \mid \mathbf{x}^{(t-1)}, \mathbf{c}^{(t)}\right) = f(x_t \mid c_t), \quad t \in \mathbb{N}.$$

The only difference in models between that of a HMM and a SSM is that the Markov process $\{C_t\}$ is continuous-valued in the latter. However, as we will discretize the state space into a sufficiently large but finite number of states, we can evaluate an approximation of the likelihood of any given SSM, exactly like the discrete-valued HMM.

The discretization procedure is as follows: For some given SSM, we consider an essential range $[b_0, b_m]$ of possible values of C_t. This range is then subdivided into m subintervals $B_i = (b_{i-1}, b_i)$, $i = 1, \ldots, m$. These subintervals need not be on a equidistant grid, however, for simplicity and computational ease, we assume they are equidistant. As such, they are all of the length $h = (b_m - b_0)/m$. Denote b_i^* as a representative point in B_i, for example the midpoint. By making use of the SSM dependence structure and repeatedly approximating integrals $\int_a^b f(c)dc$ by simple expressions of the form $(b-a)f(c^*)$, the likelihood of the observations $\mathbf{x}^{(T)}$ can be approximated

[6] We return to indexing at $t = 1$ to align with the cited literature and to bind the autoregressive theory with the state-space theory.

as

$$
\begin{aligned}
\mathcal{L}_T &= \underbrace{\int \dots \int}_{T-\text{integrals}} f_{\mathbf{X}^{(T)},\mathbf{C}^{(T)}}\left(\mathbf{x}^{(T)}, \mathbf{c}^{(T)}\right) \mathrm{d}c_T \dots \mathrm{d}c_1 \\
&\overset{\dagger}{=} \int \dots \int f_{\mathbf{X}^{(T)}|\mathbf{C}^{(T)}}\left(\mathbf{x}^{(T)} \mid \mathbf{c}^{(T)}\right) f_{\mathbf{C}^{(T)}}\left(\mathbf{c}^{(T)}\right) \mathrm{d}c_T \dots \mathrm{d}c_1 \\
&\overset{\dagger\dagger}{=} \int \dots \int f_{C_1}(c_1) f_{X_1|C_1}(x_1 \mid c_1) \prod_{t=2}^{T} f_{C_t|C_{t-1}}(c_t \mid c_{t-1}) f_{X_t|C_t}(x_t \mid c_t) \mathrm{d}c_T \dots \mathrm{d}c_1 \qquad (30) \\
&\overset{\dagger\dagger\dagger}{\approx} \int_{b_0}^{b_m} \dots \int_{b_0}^{b_m} f_{C_1}(c_1) f_{X_1|C_1}(x_1 \mid c_1) \prod_{t=2}^{T} f_{C_t|C_{t-1}}(c_t \mid c_{t-1}) f_{X_t|C_t}(x_t \mid c_t) \mathrm{d}c_T \dots \mathrm{d}c_1 \\
&\overset{\dagger\dagger\dagger\dagger}{\approx} h^T \sum_{i_1=1}^{m} \dots \sum_{i_T=1}^{m} f_{B_{i_1}}\left(b_{i_1}^*\right) f_{X_1|B_{i_1}}(x_1 \mid b_{i_1}^*) \prod_{t=2}^{T} f_{B_{i_t}|B_{i_{t-1}}}\left(b_{i_t}^* \mid b_{i_{t-1}}^*\right) f_{X_t|B_{i_t}}\left(x_t \mid b_{i_t}^*\right).
\end{aligned}
$$

$\dagger$ follows from definition of a joint density, $\dagger\dagger$ is a notational rewriting and $\dagger\dagger\dagger$ is splitting the integrals into the range $[b_0, b_m]$. Finally, $\dagger\dagger\dagger\dagger$ is an approximation,where the innermost integral has been approximated as

$$
\int_{b_0}^{b_m} f_{C_T|C_{T-1}}(c_T \mid c_{T-1}) f_{X_T|C_T}(x_T \mid c_T) \mathrm{d}c_T \approx h \sum_{i_T}^{m} f_{B_{i_T}|C_{T-1}}(b_{i_T}^* \mid c_{T-1}) f_{X_T|B_{i_T}}\left(x_T \mid b_{i_T}^*\right).
$$

The terms appearing in the last approximation in Equation 30 are simple. However, the likelihood cannot be evaluated because of the m^T number of summands. The likelihood with a discrete state space does yields a convenient form which allows us to employ our previously developed technique of using the forward algorithm to evaluate the likelihood.

Evaluation of the Approximate Likelihood The discretization of the state space into some large number of intervals m corresponds to an approximation of the SSM by an m-state HMM. However, it is now possible to specify the components of this approximating HMM with ease.

First, Consider the initial distribution of the state process. To obtain the exact expressions given in the last line of Equation 30, we define the i'th component of the m-dimensional vector $\boldsymbol{\delta}$ to be $\delta_i = hf(b_i^*)$. then δ_i is the approximate probability of the state process falling in the interval B_i at time-1 (as it is the initial distribution). For example, assume that the state process is in its stationary distribution at the time of the first observation. Then $f(b_i^*)$ is the density of the normal distribution evaluated at b_i^* with $\mathbb{E}[C_t^*] = 0$ an $\mathbb{V}[C_t^*] = \sigma_\varepsilon^2 / (1 - \rho^2)$. In the exact same manner, define an $m \times m$ t.p.m $\boldsymbol{\Gamma} = (\gamma_{ij})_{i,j=1}^m$ by specifying $\gamma_{ij} = hf(b_j^* \mid b_i^*)$. The transition probabilities γ_{ij} are the approximate probability of the value of the state process falling into the intervals B_j at time t given that the process is in interval B_i at time $t-1$. For the Gaussian AR(1) state process, the values of γ_{ij} is h times the density of the normal distribution with mean ρb_i^* and variance σ_ε^2

evaluated at b_j^*. Lastly, we define the component $\boldsymbol{P}(x_t)$ to be the $m \times m$ diagonal matrix with ith entry corresponding to $f(x_t \mid b_i^*)$. This corresponds to an approximation of the conditional density of x_t given that the state process takes some value in the interval B_i at time t.

Assembling the components just defined, we can rewrite the multiple-sum expression for the apprximate likelihood given in Equation 30 in the form of a matrix product

$$h^T \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} f_{B_{i_1}}(b_{i_1}^*) f_{X_1|B_{i_1}}(x_1 \mid b_{i_1}^*) \prod_{t=2}^{T} f_{B_{i_t}|B_{i_{t-1}}}(b_{i_t}^* \mid b_{i_{t-1}}^*) f_{X_t|B_{i_t}}(x_t \mid b_{i_t}^*)$$
$$= \boldsymbol{\delta} \boldsymbol{P}(x_1) \boldsymbol{\Gamma} \boldsymbol{P}(x_2) \boldsymbol{\Gamma} \boldsymbol{P}(x_3) \cdots \boldsymbol{\Gamma} \boldsymbol{P}(x_{T-1}) \boldsymbol{\Gamma} \boldsymbol{P}(x_T) \mathbf{1}_N^\top.$$

Estimation Issues and Assessment According to [60, p. 160], numerical maximization of the likelihood given in Equation 30 is feasible even when the observation count and m is fairly large, which is what would be required for a relatively close approximation to the likelihood. In general, it seems to be that values around $m = 50$ stabilize [60, 29]. Furthermore, the number of parameters does not depend on the magnitude of m. The entries of the approximate tpm $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times m}$, depends only on state process parameters of the SSM. The range $[b_0, b_m]$ has to be chosen such that the range is sufficiently large to cover the essential range of the state process. However, if the range is chosen too large for a fixed m the grid becomes too coarse, degrading the accuracy of the likelihood approximation.

Complications remain the same as for the N-state HMM, as we are essentially fitting a m-state hidden Markov model; Local maxima, parameter constraints and especially, numerical under- and overflow. The techniques described for the HMM to handle the estimation complication will be used in an identical manner.

As for the HMM, we extract the numerically estimated Hessian of the log-likelihood for the estimated parameters using the base `R` function `nlm`. Furthermore, we can use the Viterbi algorithm for decoding, pseudo-residuals and forecasts, exactly as for the HMM.

The BS-SSM$_\beta$ specified in (28) is parameterized by $\boldsymbol{\zeta} = (\rho, \sigma_\varepsilon, \mu_0, \sigma, \beta_\mu, \beta_\sigma)$, where $(\rho, \sigma_\varepsilon)$ govern the latent AR(1) factor $\{C_t\}$ and $(\mu_0, \sigma, \beta_\mu, \beta_\sigma)$ determine the time-varying drift μ_t and volatility σ_t of returns. This parameterization is not globally identifiable from the return series $\{X_t\}$ alone. The source of non-identifiability is a scale (and sign) invariance in the latent factor and its loadings.

Specifically, for any $a > 0$ define a rescaled latent factor $C_t' = a\, C_t$ and a corresponding reparameterization

$$\rho' = \rho, \qquad \mu_0' = \mu_0, \qquad \sigma' = \sigma,$$
$$\sigma_\varepsilon' = a\, \sigma_\varepsilon, \qquad \beta_\mu' = \beta_\mu / a, \qquad \beta_\sigma' = \beta_\sigma / a.$$

Under this mapping, the state equation preserves its form,

$$C'_t = \rho\, C'_{t-1} + \varepsilon'_t, \qquad \varepsilon'_t \sim \mathcal{N}\big(0, (\sigma'_\varepsilon)^2\big)\,,$$

and the implied drift and volatility processes are unchanged:

$$\begin{aligned}\mu'_t &= \mu'_0 + \beta'_\mu C'_t = \mu_0 + \beta_\mu C_t = \mu_t,\\ \sigma'_t &= \sigma' \exp(\beta'_\sigma C'_t) = \sigma \exp(\beta_\sigma C_t) = \sigma_t.\end{aligned}$$

Consequently, the conditional distribution of returns $X_t \mid C_t$ is unchanged and therefore the induced law of the observed return sequence $\{X_t\}$ is identical under $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}'$. In addition, there is a sign invariance obtained by mapping $C'_t = -C_t$ and $(\beta'_\mu, \beta'_\sigma) = (-\beta_\mu, -\beta_\sigma)$ (with $\rho, \sigma_\varepsilon, \mu_0, \sigma$ unchanged), which again leaves (μ_t, σ_t) and the distribution of $\{X_t\}$ unchanged. As a result, the individual parameters $(\sigma_\varepsilon, \beta_\mu, \beta_\sigma)$ and the absolute scale and orientation of C_t are not uniquely identified from returns alone.

This lack of global identifiability does not affect the quantities used in the empirical analysis. The maximized log-likelihood depends only on the induced distribution of $\{X_t\}$ and is invariant along the above reparameterization ridge, so AIC and BIC comparisons between the BSM, BS-HMM and BS-SSM$_\beta$ are well defined. Likewise, conditional forecast distributions $F_{X_{t+h}|\mathcal{F}_t}$, point forecasts and forecast intervals depend on the implied dynamics of (μ_t, σ_t) and are therefore unchanged by any scale or sign transformation that preserves (μ_t, σ_t). The same applies to pseudo-residuals, which are constructed from the fitted conditional CDF of X_t and hence depend on the return distribution implied by the model rather than on a particular normalization of the latent factor. All likelihood-based and distribution-based diagnostics reported here (AIC/BIC, pseudo-residuals and forecast-error measures) are therefore invariant to this non-identifiability.

Some components of $\boldsymbol{\zeta}$ remain identifiable in the usual sense. The persistence parameter ρ and the baseline parameters (μ_0, σ) are not affected by the above invariances and are pinned down by the return dynamics up to the usual sampling variation and numerical optimization considerations. What is not uniquely identified is the internal decomposition of time variation in (μ_t, σ_t) into the scale of the latent factor (through σ_ε) versus the strength of its loadings $(\beta_\mu, \beta_\sigma)$. In this thesis, C_t is therefore interpreted as a relative latent index of the return environment over time rather than a factor with an intrinsically meaningful absolute scale. Statements about time-variation and ordering of regimes are invariant and can be interpreted, whereas structural claims about the absolute magnitude of $(\sigma_\varepsilon, \beta_\mu, \beta_\sigma)$ are not.

Simulation Firstly, we simulate the BS-SSM in Equation 27, that is, a Black–Scholes model with an AR(1) latent state $C_t = \rho C_{t-1} + \varepsilon_t$ with innovations $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ that multiplicatively scales both drift and volatility in the log-return formulation X_t (cf. Equation 4). The baseline parameters are $\mu = 0.05$, $\sigma = 0.15$, $\rho = 0.98$ (high persistence), $\sigma_\varepsilon = 0.10$, with $S_0 = 100$, $n = 25000$ and daily observations $\Delta = 1/252$ (approximately $25000/252 \approx 99$ years). The simulated price path is shown in Figure 12 and estimated versus true parameters are reported in Table 3. We use `nlm` in `R` to maximise the discretised likelihood in Equation 30 for the BS-SSM.

We then consider the BS-SSM$_\beta$ specification in Equation 28, where the same AR(1) latent state enters the drift and log-volatility via linear factor loadings, $\mu_t = \mu_0 + \beta_\mu C_t$ and $\sigma_t = \sigma \exp(\beta_\sigma C_t)$. For the simulation we choose $\mu_0 = 0.05$, $\beta_\mu = 0.20$, $\sigma = 0.15$, $\beta_\sigma = 0.60$, together with $\rho = 0.98$, $\sigma_\varepsilon = 0.10$, $S_0 = 100$, $n = 25000$ and $\Delta = 1/252$. The corresponding price path is shown in Figure 13 and Table 4 collects true and estimated parameters.

In both cases, the relative estimation errors for are small and do not raise concerns, bearing in mind the approximation error from the Euler discretization and grid-based likelihood. The only stand-out parameter estimate is the loading factor β_μ in the BS-SSM$_\beta$ that is somewhat less precisely estimated. This is consistent with expectation given that it primarily affects the conditional mean rather than the volatility, which we know to be prone to erroneous estimation.

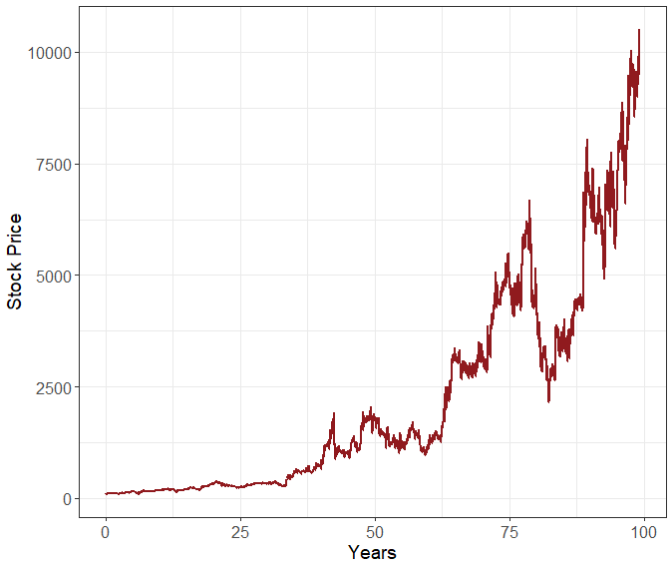


Figure 12: *BS-SSM simulated stock price path.*

Parameter	True Value	Estimated Value	Relative Error (%)
ρ	0.98000	0.98039	0.04
σ_ε	0.10000	0.09967	0.33
μ	0.05000	0.05062	1.24
σ	0.15000	0.14653	2.31

Table 3: *True vs. estimated parameters for the BS-SSM, including relative estimation errors.*

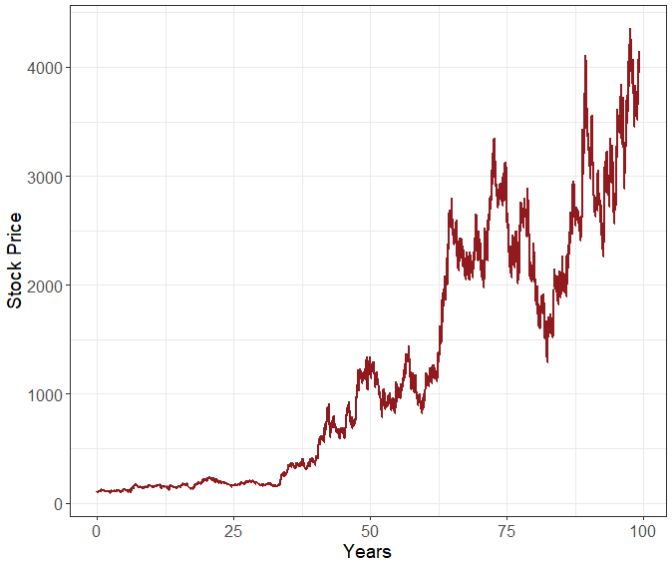


Figure 13: *BS-SSM$_\beta$ simulated stock price path.*

Parameter	True Value	Estimated Value	Relative Error (%)
ρ	0.98000	0.98082	0.08
σ_ε	0.10000	0.10006	0.06
μ_0	0.05000	0.05062	1.24
σ	0.15000	0.14814	1.24
β_μ	0.20000	0.22590	12.95
β_σ	0.60000	0.59251	1.25

Table 4: *True vs. estimated parameters for the BS-SSM$_\beta$, including relative estimation errors.*

2.4 Model Selection Criteria & Assessment

2.4.1 Information Criteria: AIC & BIC

Two of the most popular approaches to model selection for HMMs will be used: The Akaike Information Criterion (AIC) and The Bayesian Information Criterion (BIC). These are supplementary methods to those discussed previously.

Assume that $x_1, \ldots, x_T$ were generated by the true data generating process, f and that one is interested in determining which model to choose among two different approximating families $\{g_1 \in \mathcal{G}_1\}$ and $\{g_2 \in \mathcal{G}_2\}$ under some criteria of being "the best". We thus need some operator to determine the lack of fit between the true data generating model and the fitted models, $\Delta(f, \hat{g}_1)$ and $\Delta(f, \hat{g}_2)$. A immediate issue that arises is the lack of knowledge of f. As such, we can not determine from this discrepancy which model to select. However, we can use model selection criteria, $\widehat{\mathbb{E}}_f\big[\Delta(f, \hat{g}_1)\big]$ and $\widehat{\mathbb{E}}_f\big[\Delta(f, \hat{g}_2)\big]$. These quantities bases selection on estimators of the expected discrepancies. The model selection criterion simplifies to the Akaike information criterion [60, p. 98] which, briefly stated, arises of the Kullback–Leibler discrepancy and conditions listed in [31, Appendix A]:

$$\text{AIC} = \underbrace{-2 \log \mathcal{L}_T}_{\text{measure of fit}} + \underbrace{2p}_{\text{penalty}}, \tag{31}$$

where $\mathcal{L}$ is the log-likelihood of the fitted model and p denotes the number of parameters of the model[7]. It is immediately clear that increasing the number of parameters, by increasing the number of states or state-dependent parameters, will penalize the AIC. To compare model performances in terms of AIC, we follow [8, pp. 270-272] to some degree; Let Δi denote the difference in AIC between the best model (i.e. smallest AIC) and the one of comparison. The rule of thumb then states that we can assess the relative merits of models by:

- $\Delta i \leq 2 \Rightarrow$ Substantial support (evidence).
- $4 \leq \Delta i \leq 7 \Rightarrow$ Considerably less support (evidence).
- $\Delta i > 10 \Rightarrow$ Essentially no support (evidence).

Note, that [9] relaxed the rule of thumb and thus $2 \leq \Delta i \leq 7$ have some support and should seldom be disregarded. However, this is not sufficient for model assessment as discussed in Section 2.2.7.

Another approach to model selection is the Bayesian philosopy. The Bayesian philosophy to model selection differs slightly to the AIC approach. The Bayesian philosophy is to select the family which is estimated to be most likely to be true. Consistent with the Bayesian paradigm, in the first step before considering observations at hand, one specifies the prior probabilities, that

[7] see Section 2.2.7 for number of parameter determination.

f stems from the approximating families $\mathcal{G}_1, \mathcal{G}_2$, namely, $\mathbb{P}(f \in \mathcal{G}_1)$ and $\mathbb{P}(f \in \mathcal{G}_2)$. Secondly, one computes and compares the posterier probabilities that f belongs to the approximating families given the observations, namely, $\mathbb{P}\left(f \in \mathcal{G}_1 \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right)$ and $\mathbb{P}\left(f \in \mathcal{G}_2 \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right)$. Briefly stated, under conditions seen in [58], the Bayesian information criterion arises [60, p. 98]:

$$\text{BIC} = \underbrace{-2 \log \mathcal{L}_T}_{\text{measure of fit}} + \underbrace{p \log T}_{\text{penalty}}, \tag{32}$$

where $\mathcal{L}_T$ and p are as for the AIC and T is the number of observations, which is obviously not present whatsoever for the AIC. Compared to the AIC, the penalty term of the BIC has more weight for $T > e^2$, which holds in most practical applications. Thus, the BIC does, in general, favour models with fewer parameters than the AIC.

Summarizing, in both cases, the best model in the family is the one that minimizes these information criteria. Clearly, AIC does not depend directly on the sample size, T. Moreover, AIC presents the danger that it might overfit, whereas BIC presents the danger that it might underfit, simply in virtue of how each criterion penalize free parameters (see the under-braced penalty-terms in Equation 31 and Equation 32).

2.4.2 Pseudo-Residuals

Even after selecting what seems to be the best model according to some chosen criterion, it is still necessary to determine whether the model actually provides a good fit to the data. This requires tools that can assess the overall adequacy of the model and help identify potential outliers. In classical settings such as normal-theory regression, residuals are a well-known and widely used method for checking model fit. In this section, we introduce quantities called pseudo-residuals (also referred to as quantile residuals), which extend this idea to more general models and serve a similar purpose in the context of HMMs. We present two types of pseudo-residuals, both of which rely on the ability to compute likelihoods efficiently, something that HMMs naturally allow. The theory presented is based on that of [15, pp. 236-244] which they note is a special case of Cox–Snell residuals [12].

Each X_t has a distribution that depends on some latent state in the state space. As such, assessing outliers or model fit is non-trivial, since the conditional distribution of each X_t changes over time and depends on the hidden state sequence. A commonly used approach in HMMs for assessment of model fit is to transform the observations to a common scale using pseudo-residuals $\{z_t\}_{t=1}^{T}$, constructed via the probability integral transform [60, pp. 101-106]:

I. Transform a observation x_t to $u_t = F_{X_t}(x_t) \sim \mathcal{U}[0,1]$, where F_{X_t} is the CDF of X_t.

II. Transform u_t to $z_t = \Phi^{-1}(u_t) \sim \mathcal{N}(0,1)$, where Φ is the standard normal CDF.

III. If the model is correctly specified, then the pseudo-residuals, $z_t = \Phi^{-1}\left(F_{X_t}(x_t)\right)$, should be approximately independent and standard normally distributed. These can be evaluated using histograms and Q–Q plots.

We show the properties in **I.** and **II.** to be the case.

Proposition 2.7. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a real-valued r.v. with cumulative distribution function F_X. Suppose F_X is continuous and strictly increasing (hence invertible with inverse F_X^{-1}). Define*

$$U := F_X(X) \qquad \textit{and} \qquad Z := \Phi^{-1}(U),$$

where Φ is the standard normal distribution function. Then $U \sim \mathcal{U}[0,1]$ and $Z \sim \mathcal{N}(0,1)$.

Proof. First, for any $u \in [0,1]$,

$$\begin{aligned}
\mathbb{P}(U \leq u) &= \mathbb{P}\left(F_X(X) \leq u\right) \\
&\overset{\dagger}{=} \mathbb{P}\left(X \leq F_X^{-1}(u)\right) \\
&= F_X\left(F_X^{-1}(u)\right) \\
&= u,
\end{aligned}$$

where † follows since F_X is strictly increasing. Next, for any $z \in \mathbb{R}$,

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \leq z) \\
&= \mathbb{P}\left(\Phi^{-1}(U) \leq z\right) \\
&= \mathbb{P}\left(U \leq \Phi(z)\right) \\
&= F_U\left(\Phi(z)\right) \\
&= \Phi(z).
\end{aligned}$$

Therefore $Z \sim \mathcal{N}(0,1)$. □

Ordinary Pseudo-Residuals The first approach assesses individual observations by flagging values that appear unusually extreme relative to the fitted model and the remainder of the series, suggesting that they may be atypical in nature or origin. Operationally, this is done by constructing pseudo-residuals $\{z_t\}_{t=1}^T$ from the conditional distribution of $X_t \mid \mathbf{X}^{(-t)}$. For continuous observations, the (Gaussian) ordinary pseudo-residual is defined as

$$z_t = \Phi^{-1}\left(\mathbb{P}\left(X_t \leq x_t \mid \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}\right)\right). \tag{33}$$

If the model is correctly specified, z_t is approximately a realization of a standard normal random variable. Using results from Section 2.2.6, specifically by integrating over Equation 22 and Equation 23, the ordinary pseudo-residuals can be computed. In the empirical assessment, however, we place greater emphasis on forecast pseudo-residuals, since our primary interest is not in idiosyncratic outliers relative to the full-sample fit, but rather in the model's one-step predictive performance.

Forecast Pseudo-Residuals The second approach targets observations that appear unusually extreme relative to the model's predictions based solely on information available prior to time t. The key object is therefore the conditional distribution of X_t given $\mathbf{X}^{(t-1)}$. For continuous observations, the corresponding pseudo-residual is

$$z_t = \Phi^{-1}\left(\underbrace{\mathbb{P}\left(X_t \leq x_t \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right)}_{(*)}\right). \tag{34}$$

To compute the forecast pseudo-residuals, we evaluate the one-step-ahead forecast distribution of X_t given the observed history up to time $t-1$, i.e. $(*)$ in Equation 34. Note that

$$\begin{aligned}
\mathbb{P}\left(X_t \leq x_t \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right) &= \sum_{j\in\mathcal{C}} \mathbb{P}\left(X_t \leq x_t, C_t = j \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right) \\
&\overset{\dagger}{=} \sum_{j\in\mathcal{C}} \underbrace{\mathbb{P}\left(X_t \leq x_t \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}, C_t = j\right)}_{:=F_{X_t,j}(x_t)} \underbrace{\mathbb{P}\left(C_t = j \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right)}_{:=\psi_t(j)} \\
&\overset{\dagger\dagger}{=} \sum_{j\in\mathcal{C}} \psi_t(j) F_{X_t,j}(x_t),
\end{aligned}$$

where $\dagger$ follows from the Law of Total Probability and $\dagger\dagger$ from the HMM conditional-independence assumption. Thus, in the HMM setting, the one-step-ahead forecast distribution is a mixture of state-dependent conditional distributions, with $\psi_t(j) = \mathbb{P}\left(C_t = j \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right)$ denoting the one-step-ahead predicted state probabilities and for $q, p \in \mathbb{R}$

$$F_{X_t,j}(x_t) = \Phi\left(\frac{x_t - m_j}{s_j}\right),$$

with $m_j := \left(\mu_j - \frac{1}{2}\sigma_j^2\right)\Delta$ and $s_j^2 = \sigma_j^2\Delta$.

The predicted state probabilities are obtained by propagating the filtered probabilities forward

using the transition probabilities and the normalized vector of forward variables:

$$\begin{aligned}
\psi_t(j) &= \mathbb{P}\left(C_t = j \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right) \\
&= \sum_{i \in \mathcal{C}} \underbrace{\mathbb{P}\left(C_t = j \mid C_{t-1} = i\right)}_{:=\gamma_{i,j}} \underbrace{\mathbb{P}\left(C_{t-1} = i \mid \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right)}_{:=\phi_{t-1}(i)} \\
&= \sum_{i \in \mathcal{C}} \gamma_{ij} \phi_{t-1}(i) \\
&\Rightarrow \\
\boldsymbol{\psi}_t &= \boldsymbol{\phi}_{t-1} \boldsymbol{\Gamma}.
\end{aligned}$$

For numerical stability, we implement these quantities using exponentiated, normalized forward probability vectors.

In the BS-HMM, the state-dependent cumulative distributions $F_{X_t,j}(x_t)$ are governed by the Gaussian specification

$$X_t \mid \{C_t = i\} \sim \mathcal{N}\left(\left(\mu_i - \frac{1}{2}\sigma_i^2\right)\Delta, \sigma_i^2 \Delta\right).$$

The pseudo-residuals are then computed as

$$z_t = \Phi^{-1}\left(\sum_{j \in \mathcal{C}} \psi_t(j) \cdot F_{X_t,j}(x_t)\right), \quad t = 2, \ldots, T.$$

For the first residual z_1, the one-step-ahead state probabilities $\psi_1(j)$ cannot be propagated from previous forward probabilities, since there is no observation prior to $X_1 = x_1$. A convenient circumvention is to approximate them using the stationary distribution $\boldsymbol{\delta}$, which represents the long-run state probabilities of the underlying Markov chain. Accordingly, the first residual is computed as

$$z_1 = \Phi^{-1}\left(\sum_{j \in \mathcal{C}} \delta_j \cdot F_{X_t,j}(x_t)\right).$$

If the model is correctly specified, the pseudo-residuals $\{z_t\}_{t=1}^T$ should be approximately independent and standard normally distributed.

3 Data (II of II)

Throughout the empirical analysis we work with daily log-returns rather than price levels. Let S_t denote the closing level of the S&P 500 on trading day t. For $t = 2, \ldots, T$, the one-day log-return is

$$x_t \;=\; \log S_t - \log S_{t-1} \;=\; \log\!\left(\frac{S_t}{S_{t-1}}\right). \quad (35)$$

This transformation is standard in financial econometrics for three main reasons.

First, broad equity price levels are non-stationary over long samples, whereas returns are typically closer to weak stationarity. Since likelihood-based time-series methods (e.g. ARMA, GARCH and state-space models) are formulated under stationarity assumptions, working with returns mitigates spurious dynamics induced by trending levels and yields a series with a stable mean close to zero and finite variance (though with pronounced conditional heteroskedasticity).

Second, log-returns aggregate additively across time. For any horizon $h \in \mathbb{N}$,

$$x_{t:t+h} \;:=\; \log\!\left(\frac{S_{t+h}}{S_t}\right) \;=\; \sum_{k=t+1}^{t+h} x_k \quad (36)$$

By contrast, simple (arithmetic) returns $r_t := (S_t - S_{t-1})/S_{t-1}$ compound multiplicatively. For high-frequency horizons where $|r_t|$ is small, $\log(1+r_t) \approx r_t$ which follows from $\log(1+r) = r - \frac{1}{2}r^2 + \mathcal{O}(r^3)$. As such, log and simple returns are approximate, while Equation 36 holds exactly.

Third, log-returns are invariant to index rebasings: if prices are rescaled by any constant $c > 0$, then $\log(cS_t) - \log(cS_{t-1}) = x_t$. This facilitates comparisons across periods and improves numerical stability relative to working with raw index levels.

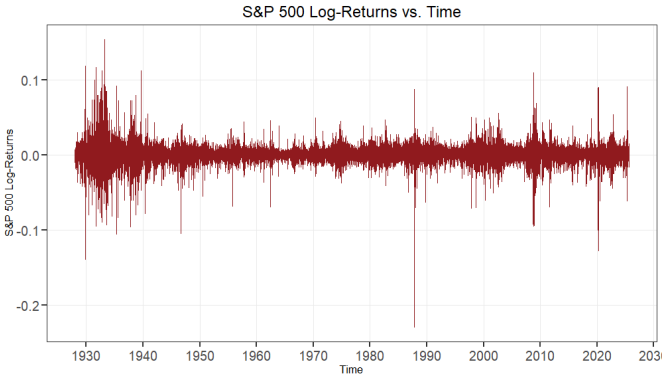


Figure 14: *S&P 500 time series of daily log-returns.*

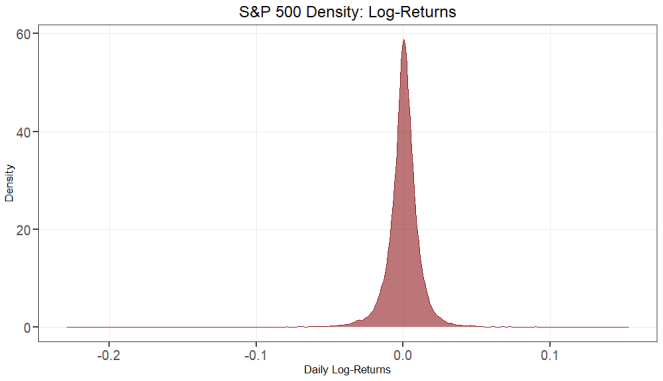


Figure 15: *Empirical density of S&P 500 daily log-returns.*

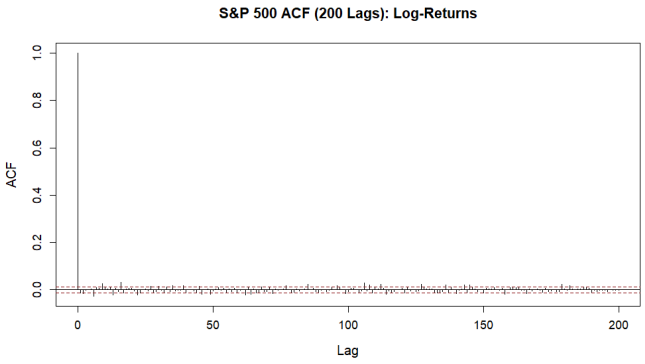


Figure 16: *Sample autocorrelation function (200 lags) for S&P 500 daily log-returns.*

To examine the transformed series, Figure 14 displays the log-returns, while Figure 15 shows their empirical distribution, which is unimodal and approximately symmetric around zero but exhibits

occasional extreme observations (tail events). For reference, the first quartile is -0.0045519, the median is 0.0004940 and the third quartile is 0.005458. The largest observation is 0.1536613 and the smallest is -0.2289973. Finally, Figure 16 reports the sample autocorrelation over 200 lags, indicating limited linear dependence in daily returns.

Asymptotic Inference and HAC Standard Errors For the global (constant-dividend) estimator, define the population mean of the log-dividend series and its sample analogue as

$$\alpha := \mathbb{E}\left[q_t^{(\log)}\right], \qquad \widehat{\alpha} := \frac{1}{T}\sum_{t=1}^{T} q_t^{(\log)},$$

and let $u_t := q_t^{(\log)} - \alpha$. We treat $\left\{q_t^{(\log)}\right\}_{t\in\mathbb{Z}}$ as a weakly dependent time series in calendar time and, for the purpose of HAC inference, assume it is (approximately) stationary and ergodic with absolutely summable auto-covariances

$$\Psi_h := \mathrm{Cov}\left(u_t, u_{t+h}\right), \qquad \sum_{h=-\infty}^{\infty} |\Psi_h| < \infty.$$

In addition, we impose a (dependent) CLT for the sample mean[8],

$$\sqrt{T}\left(\widehat{\alpha} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_q\right), \qquad \Omega_q := \Psi_0 + 2\sum_{h=1}^{\infty}\Psi_h,$$

so that $\mathbb{V}\left(\widehat{\alpha}\right) \approx \Omega_q/T$ by Lemma 2.1. We estimate Ω_q using a Newey-West HAC estimator [41], implemented via the intercept-only regression

$$q_t^{(\log)} = \alpha + u_t, \qquad \widehat{\alpha} = \arg\min_{\alpha}\sum_{t=1}^{T}\left(q_t^{(\log)} - \alpha\right)^2.$$

Let $\widehat{u}_t := q_t^{(\log)} - \widehat{\alpha}$ and define, for $h \geq 0$,

$$\widehat{\Psi}_h := \frac{1}{T}\sum_{t=h+1}^{T}\widehat{u}_t\,\widehat{u}_{t-h}.$$

With bandwidth H_T and Bartlett weights $w_h := 1 - \frac{h}{H_T+1}$, the Newey-West estimator is

$$\widehat{\Omega}_q = \widehat{\Psi}_0 + 2\sum_{h=1}^{H_T} w_h\,\widehat{\Psi}_h, \qquad \widehat{\mathrm{SE}}\left(\widehat{\alpha}\right) = \sqrt{\widehat{\Omega}_q/T}.$$

[8]For sufficient conditions ensuring such CLTs under weak dependence (specifically stationary ergodic processes with absolutely summable autocovariances), see [2, Thm. 7.7.8] or [25, Prop. 7.11].

For the state-wise estimators, note that $\mathcal{T}_i$ is generally non-contiguous in calendar time. Our implementation therefore applies HAC inference to the ordered state-occurrence subsequence. Specifically, let

$$t_{i,1} < t_{i,2} < \cdots < t_{i,n_i}$$

denote the ordered elements of $\mathcal{T}_i$ and define the event-time subsequence

$$q_{i,k} := q^{(\log)}_{t_{i,k}}, \quad k = 1, \ldots, n_i.$$

We estimate α_i from the intercept-only regression in event time,

$$q_{i,k} = \alpha_i + u_{i,k}, \qquad k = 1, \ldots, n_i,$$

so that

$$\widehat{\alpha}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} q_{i,k} = \frac{1}{n_i} \sum_{t \in \mathcal{T}_i} q^{(\log)}_t.$$

For inference, we treat $\{u_{i,k}\}_{k \in \mathbb{Z}}$ as an (approximately) weakly dependent stationary sequence in the event-time index k with auto-covariances

$$\Psi_{h,i} := \mathrm{Cov}\left(u_{i,k}, u_{i,k+h}\right), \qquad \Omega_{q,i} := \Psi_{0,i} + 2 \sum_{h=1}^{\infty} \Psi_{h,i},$$

and by imposing the corresponding CLT again,

$$\sqrt{n_i}\left(\widehat{\alpha}_i - \alpha_i\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_{q,i}\right), \qquad \mathbb{V}\left(\widehat{\alpha}_i\right) \approx \Omega_{q,i}/n_i.$$

In practice, $\Omega_{q,i}$ is unknown and is estimated by Newey-West HAC on the intercept-only regression. Let $\widehat{u}_{i,k} := q_{i,k} - \widehat{\alpha}_i$ and define

$$\widehat{\Psi}_{h,i} := \frac{1}{n_i} \sum_{k=h+1}^{n_i} \widehat{u}_{i,k}\, \widehat{u}_{i,k-h}, \qquad h \geq 0.$$

With bandwidth H_{n_i} and Bartlett weights $w_h := 1 - \frac{h}{H_{n_i}+1}$, set

$$\widehat{\Omega}_{q,i} = \widehat{\Psi}_{0,i} + 2 \sum_{h=1}^{H_{n_i}} w_h\, \widehat{\Psi}_{h,i}, \qquad \widehat{\mathrm{SE}}\left(\widehat{\alpha}_i\right) = \sqrt{\widehat{\Omega}_{q,i}/n_i}.$$

Finally, annualisation is by linear rescaling,

$$\widehat{q}_i := 252\, \widehat{\alpha}_i, \qquad \widehat{\mathrm{SE}}\left(\widehat{q}_i\right) = 252\, \widehat{\mathrm{SE}}\left(\widehat{\alpha}_i\right),$$

and analogously for the global estimator $\widehat{q} := 252\,\widehat{\alpha}$.

For the state-wise HAC corrections, the lag index in $\widehat{\Psi}_{h,i}$ is an event-time lag: $h = 1$ corresponds to successive occurrences of state i in the decoded path, not to one trading day. Since the calendar-time gaps $t_{i,k} - t_{i,k-1}$ vary, the resulting HAC standard errors should be interpreted as measuring dependence across successive state-i observations rather than dependence at fixed calendar lags. Consequently, if $q_t^{(\log)}$ exhibits strong calendar-time persistence that is not well captured by event-time dependence, the state-wise HAC standard errors may under- or overstate sampling uncertainty. We adopt this event-time Newey-West procedure because it matches the implementation and provides a simple, robust correction for serial dependence within the state-specific subsequences.

Moreover, $\mathcal{T}_i$ is defined using the Viterbi-decoded states computed from estimated model parameters. The reported standard errors therefore condition on the decoded path and ignore additional uncertainty from state classification and parameter estimation. In this sense, the state-wise confidence intervals are best viewed as descriptive measures of uncertainty under a fixed-decoding approximation.

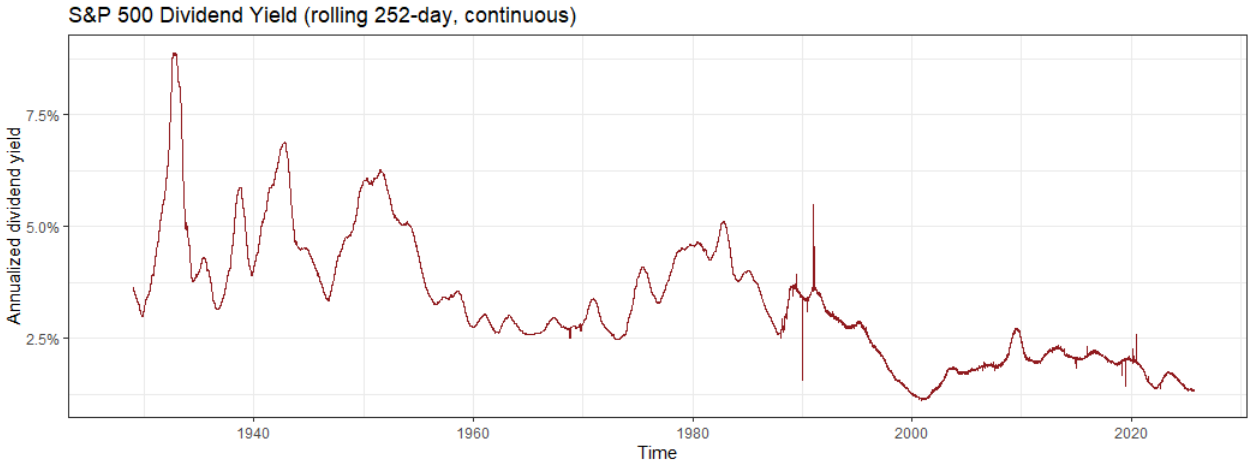


Figure 17: *Annualized dividend yield vs. time for the S&P 500, constructed from Shiller's monthly price-dividend data in the pre-TR period and from TR-PR differencing in the post-1988 period.*

We visualize the estimated dividend yields in Figure 17. Over the full sample (1927-2025), the estimated average continuous annualized dividend yield is approximately 3.3%, while over the restricted sample ending on 31 December 2019 it is approximately 3.4%. Using the HAC/Newey-West standard error, the standard error of the constant-dividend estimator $\widehat{q}$ is approximately 9.2×10^{-4} in both samples.

4 Empirical Data Application

4.1 Model Selection & Assessment

In what follows, let p denote the number of estimated parameters in a given model.

The Black-Scholes Model The information criteria for the BSM are reported in Table 5 and the residual diagnostics are shown in Figure A.3.1.

Black-Scholes Model (BSM)			
Model	$\boldsymbol{p}$	**AIC**	**BIC**
BSM	2	-139353.8	-139337.7

Table 5: *AIC and BIC for the standard BSM.*

The residuals exhibit pronounced heavy tails. This indicates that large positive and negative returns occur substantially more frequently than implied by the Gaussian assumption in the BSM and the model therefore underestimates tail risk. In particular, the time-series plot shows that the largest residuals coincide with major episodes of market stress, as documented in [20]:

- The Wall Street crash of 1929 and the recession of 1937-1938.
- The early-1980s recession and Black Monday in 1987.
- The dot-com bubble in the late 1990s and early 2000s.
- The 2008 financial crisis.

The autocorrelation structure of returns is largely unchanged and the model continues to both overestimate and underestimate returns outside periods of pronounced market stress. Overall, these results suggest that the BSM does not capture extreme economic environments adequately.

Black-Scholes Hidden Markov Models As shown in Table 6, the best-performing specification in terms of both AIC and BIC is the 5-state BS-HMM with state-dependent drift and volatility parameters (μ, σ). The second-best model is the 5-state BS-HMM with state-dependent volatility σ and state-independent drift μ. The third candidate is the 4-state BS-HMM with state-dependent (μ, σ).
Although the 5-state model with state-dependent (μ, σ) is preferred by the information criteria, it is also natural to consider the more parsimonious 4-state model with state-dependent (μ, σ) as a competing candidate, before the 5-state model with state-dependent σ only. To assess whether the additional flexibility yields implausible fits indicative of overfitting, we inspect the state-dependent fitted densities in Figure A.3.2.

Black-Scholes Hidden Markov Model (BS-HMM)					
Model	p	**AIC**	**BIC**	**ΔAIC**	**ΔBIC**
2-state BS-HMM (μ)	5	-139348.0	-139308.0	13994.0	13793.0
3-state BS-HMM (μ)	10	-139338.0	-139257.0	14004.0	13844.0
4-state BS-HMM (μ)	17	-139324.0	-139187.0	14018.0	13914.0
5-state BS-HMM (μ)	26	-139306.0	-139097.0	14036.0	14004.0
2-state BS-HMM (σ)	5	-150725.0	-150685.0	2617.0	2416.0
3-state BS-HMM (σ)	10	-152591.0	-152510.0	751.0	591.0
4-state BS-HMM (σ)	17	-153156.0	-153019.0	186.0	82.0
5-state BS-HMM (σ)	26	-153266.0	-153057.0	76.0	44.0
2-state BS-HMM (σ & μ)	6	-150746.0	-150698.0	2596.0	2403.0
3-state BS-HMM (σ & μ)	12	-152613.0	-152517.0	729.0	584.0
4-state BS-HMM (σ & μ)	20	-153199.0	-153038.0	143.0	63.0
5-state BS-HMM (σ & μ)	30	**-153342.0**	**-153101.0**	—	—

Table 6: *AIC and BIC for fitted BS-HMMs by state-dependent parameter family. ΔAIC and ΔBIC are computed relative to the globally best (lowest) AIC/BIC model; the best model therefore has "—" in place of ΔAIC and ΔBIC and is marked in bold KU-red*

Firstly, we inspect the state-dependent fitted densities in Figure A.3.2 for the best-performing specification under the information criteria, namely the 5-state BS-HMM with state-dependent (μ, σ). The densities associated with states 2 and 3 appear strikingly similar and closely resemble state 2 in the corresponding 4-state BS-HMM with state-dependent (μ, σ). This pattern may indicate that the 5-state model is effectively splitting a single regime into two highly similar regimes, which is consistent with potential overfitting. Moreover, when comparing the empirical state occupancies, the proportion of time spent in state 2 of the 4-state model is approximately equal to the combined time spent in states 2 and 3 of the 5-state model. To assess whether this similarity is also reflected in the parameter estimates, we next examine Table A.4.12.

The parameter estimates corroborate this impression. In the 5-state BS-HMM with state-dependent (μ, σ), the volatility levels in states 2 and 3 are nearly identical, with $\sigma_2 \approx \sigma_3 \approx 0.11$ (matching to the first three digits). While this magnitude is plausible in isolation, the degree of similarity raises concerns about redundancy across states. Turning to the drift parameters, μ_2 and μ_3 differ materially, which in itself is not problematic. However, both estimates are economically implausible and in particular $\mu_3 \approx -1.39$ suggests that the model is fitting extreme parameter values to accommodate specific observations. Taken together, these findings indicate that the 5-state specification yields unreasonable parameter estimates, consistent with overfitting in the present sample.

We therefore shift attention to the 4-state BS-HMM with state-dependent (μ, σ), with parameter estimates reported in Table A.4.11. The estimates appear reasonable, with no immediate outliers, aside from comparatively large volatility parameters that warrant discussion. We return to these in the next section and explain why they nonetheless admit a coherent interpretation.

Black-Scholes Continuous State-Space Models As discussed in Section 2.3.2, there is no closed-form prescription for selecting the grid range parameter $b_{\max}$ and the discretization level m; in practice, these tuning parameters are chosen based on empirical stability and numerical performance. Accordingly, we evaluate the negative log-likelihood over the grid

$$\begin{aligned}(b_{\max}, m) \in \Big\{ &(0.5, 20), (0.5, 40), (0.5, 70), (0.5, 100), (0.5, 200), \\ &(1, 20), (1, 40), (1, 70), (1, 100), (1, 200), \\ &(2, 20), (2, 40), (2, 70), (2, 100), (2, 200), \\ &(3, 20), (3, 40), (3, 70), (3, 100), (3, 200), \\ &(4, 20), (4, 40), (4, 70), (4, 100), (4, 200) \Big\},\end{aligned}$$

with the aim of identifying a specification that (i) attains a low negative log-likelihood and (ii) yields economically and numerically reasonable parameter estimates. The results are reported in Table 7 and Table 8. Entries marked by a "—" correspond to runs that fail to converge or produce clearly unreasonable parameter estimates, consistent with an overly coarse discretization.

	m				
$b_{\max}$	20	40	70	100	200
0.5	—	—	-74572.223296	-74568.761073	-74566.365696
1	—	-76247.016297	-76243.344459	-76242.462436	-76241.832368
2	—	-76635.707635	-76635.587101	-76635.558157	-76635.537345
3	—	—	**-76637.766845***	-76637.766841*	-76637.766841*
4	—	—	—	-76637.766841*	-76637.766841*

Table 7: *BS-SSM negative log-likelihoods rounded to 6 decimals; the overall minimum is in bold. Dashes indicate runs that did not converge or were too coarse. A $\star$ marks values with identical first 10 digits, indicating that they are essentially tied with the best model.*

	m				
$b_{\max}$	20	40	70	100	200
0.5	—	—	—	-76709.211244	-76709.211244
1	—	—	-76709.211207	-76709.211205	-76709.211205
2	—	—	-76709.211239	-76709.211239	-76709.211239
3	—	—	—	-76709.211243	-76709.211243
4	—	—	—	**−76709.211773***	-76709.211243

Table 8: *BS-SSM$_\beta$ negative log-likelihoods rounded to 6 decimals; the overall minimum is in bold. Dashes indicate runs that did not converge or were too coarse. Unlike the BS-SSM, the BS-SSM$_\beta$ results are all near identical for candidate models.*

From Table 7 and Table 8, the best-performing BS-SSM configuration is attained at

$$(m, b_{\max}) = (70, 3),$$

while the best-performing BS-SSM$_\beta$ configuration is attained at

$$(m, b_{\max}) = (100, 4).$$

Robustness checks across nearby grid choices indicate that the objective function has effectively stabilized around these solutions, so the model comparison is not sensitive to small perturbations of the grid. Information-criterion comparisons for the selected BS-SSM specifications are given in Table 9.

Black-Scholes State-Space Models (BS-SSM)					
Model	$\boldsymbol{p}$	**AIC**	**BIC**	**ΔAIC**	**ΔBIC**
BS-SSM	4	-153267.5	-153235.1	138.9	122.7
BS-SSM$_\beta$	6	**-153406.4**	**-153357.8**	—	—

Table 9: *AIC and BIC for the BS-SSM and BS-SSM$_\beta$. ΔAIC and ΔBIC are computed relative to the best (lowest) AIC/BIC model in the table.*

The BS-SSM$_\beta$ is clearly preferred in terms of both AIC and BIC. We therefore proceed to model assessment via pseudo-residual diagnostics, reported in Figure A.3.7 for the BS-SSM and Figure A.3.8 for the BS-SSM$_\beta$. We begin with the BS-SSM. The Q-Q-plot indicates heavy tails and negative skewness, with a particularly pronounced left tail relative to the Gaussian benchmark. This implies that large negative residuals occur more frequently and with greater magnitude than the model predicts, so the specification systematically understates downside movements and violates the normality assumption imposed on the innovations. Qualitatively, this pattern mirrors the pseudo-residual behaviour under the BSM. Moreover, the largest residuals occur on the same dates as under the BSM, suggesting that the model continues to fit poorly during extreme shifts in economic and financial conditions. A key difference, however, is the overall scale and direction of the residuals: the BS-SSM implies predicted returns that are systematically too high, whereas the BSM residuals are more nearly symmetric around zero. In particular, pseudo-residuals with magnitudes exceeding 5 in absolute value occur frequently. Such behavior indicates substantial model misspecification; consequently, inference and economic conclusions drawn from this specification should be treated with considerable caution. We return to an intuitive discussion of why the BS-SSM is not well-suited for modeling equity prices in Section 5.

In contrast, the BS-SSM$_\beta$ still exhibits a heavy left tail, but the diagnostics improve markedly relative to the BS-SSM. The pseudo-residuals are generally of a reasonable magnitude, with remaining outliers concentrated in the same extreme periods noted above. On this basis, we select the BS-SSM$_\beta$ as the specification used for presentation in the remainder of the empirical analysis.

4.2 Model Presentation

Black-Scholes Model First, we present the baseline BSM fitted to daily log-returns on the S&P 500 price index over the baseline sample period (see Section 1 and Section 3). Parameter estimates are reported in Table 10. In Figure 18, we overlay the fitted BSM density on the empirical distribution of daily log-returns using these estimates. All parameters are reported in annualized units. Standard errors and 95% confidence intervals are obtained from the inverse Hessian of the minimized negative log-likelihood.

The point estimate of the dividend yield is $\widehat{q} = 0.0330$ with a tight 95% confidence interval constructed using HAC standard errors, [0.0321, 0.0339], corresponding to an annual dividend yield of approximately 3.3%. The capital-gain drift of the price index is estimated as $\widehat{\mu}_{\text{cap}} = 0.0745$, with a 95% confidence interval [0.0360, 0.1131] that lies entirely above zero. This indicates a statistically and economically significant positive expected excess return on the index. Adding the dividend component, the implied total-return drift is $\widehat{\mu}_{\text{tot}} = 0.1075$, corresponding to an expected annual total return of about 10.8%.

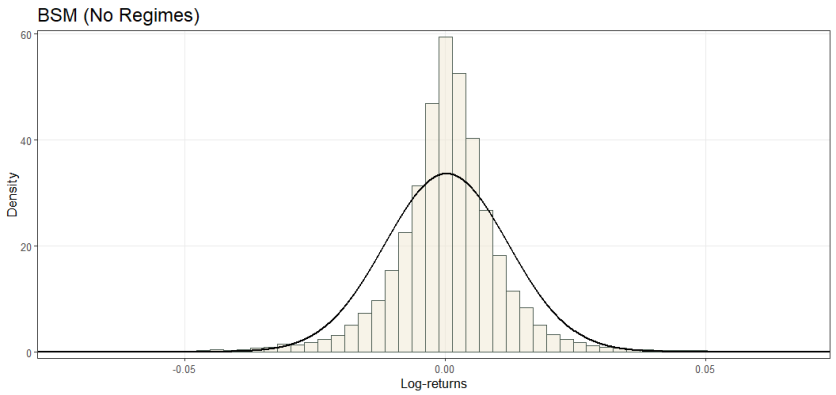


Figure 18: *Histogram of daily log-returns with the fitted BSM density overlaid.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{q}$	0.0330 (0.0321, 0.0339)	0.0009220
$\widehat{\mu}_{\text{cap}}$	0.0745 (0.0360, 0.1131)	0.01967
$\widehat{\mu}_{\text{tot}}$	0.1075 (—, —)	—
$\widehat{\sigma}$	0.1883 (0.1866, 0.1900)	0.0008761

Table 10: *BSM parameter estimates with 95% confidence intervals based on the inverse Hessian of the minimized negative log-likelihood.*

The volatility estimate is $\widehat{\sigma} = 0.1883$ with a narrow 95% confidence interval [0.1866, 0.1900], indicating that unconditional return volatility is tightly identified by the sample. Interpreted in annual terms, the model implies an S&P 500 volatility of approximately 18.8%. Overall, the BSM benchmark corresponds to a stable annual dividend yield around 3%, an expected total return around 11% and a volatility just below 20%. We use these constant-parameter estimates as a reference point when assessing the additional flexibility and empirical performance of the regime-switching specifications developed in the subsequent analysis.

Black–Scholes Hidden Markov Model We now present the 4-state BS-HMM with state-dependent drift and volatility (μ_i, σ_i). For ease of interpretation, the main parameter estimates are reported in Table A.4.11 and, in a more readable format, in Table 11. The remaining parameter estimates are reported in Appendix A.4 for the 11 BS-HMMs. The MLEs suggest that the four latent states correspond to distinct market regimes that can be interpreted as two expansionary (bull) states and two contractionary (bear) states of differing severity. States 1 and 2 are associated with positive expected capital gains, whereas states 3 and 4 imply negative expected capital gains. Across states, annualised drift and volatility differ substantially, ranging from a tranquil expansion regime to a pronounced crisis regime. The estimated dividend yields q_i are similar across the first three states and increase materially only in the most adverse regime.

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.2346 (0.1990, 0.2703)	0.01820
$\widehat{\mu}_{\text{cap},2}$	0.1062 (0.0621, 0.1502)	0.02246
$\widehat{\mu}_{\text{cap},3}$	-0.1031 (-0.2193, 0.0132)	0.05930
$\widehat{\mu}_{\text{cap},4}$	-0.3818 (-0.9260, 0.1623)	0.27763
$\widehat{q}_1$	0.0310 (0.0289, 0.0330)	0.00104
$\widehat{q}_2$	0.0348 (0.0329, 0.0368)	0.00099
$\widehat{q}_3$	0.0321 (0.0293, 0.0349)	0.00145
$\widehat{q}_4$	0.0512 (0.0426, 0.0599)	0.00442
$\widehat{\mu}_{\text{tot},1}$	0.2656 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1410 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	-0.0710 (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	-0.3306 (—, —)	—
$\widehat{\sigma}_1$	0.0707 (0.0680, 0.0734)	0.00139
$\widehat{\sigma}_2$	0.1270 (0.1229, 0.1311)	0.00210
$\widehat{\sigma}_3$	0.2301 (0.2205, 0.2396)	0.00487
$\widehat{\sigma}_4$	0.5651 (0.5329, 0.5974)	0.01646
$\widehat{\mathbf{\Gamma}} =$	$\begin{pmatrix} 0.9649\ (0.0044) & 0.0341\ (0.0045) & 0.0002\ (0.0006) & 0.0007\ (0.0005) \\ 0.0207\ (0.0030) & 0.9683\ (0.0033) & 0.0110\ (0.0015) & 0.0000\ (—) \\ 0.0000\ (0.0000) & 0.0259\ (0.0033) & 0.9632\ (0.0039) & 0.0109\ (0.0021) \\ 0.0000\ (0.0000) & 0.0000\ (0.0000) & 0.0517\ (0.0094) & 0.9483\ (0.0094) \end{pmatrix}$	
$\widehat{\boldsymbol{\delta}} =$	$(0.2766\ (0.0254), 0.4675\ (0.0233), 0.2080\ (0.0220), 0.0479\ (0.0099))$	

Table 11: *4-state BS-HMM with state-dependent drift $\mu_{cap,i}$ and volatility σ_i. One standard error for $\widehat{\mathbf{\Gamma}}$ is not reported because the inverse Hessian is numerically ill-conditioned at the optimum, yielding a non-positive delta-method variance for that entry; it is therefore marked by —. State-dependent parameter estimates are reported with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale via the delta method. In each state, $\mu_{tot,i} = \mu_{cap,i} + q_i$. The bottom block reports $\widehat{\mathbf{\Gamma}}$ and $\widehat{\boldsymbol{\delta}}$.*

State 1 is characterised by a high annualised capital-gains drift of $\widehat{\mu}_{\text{cap},1} = 0.2346$, corresponding to an expected annual price appreciation of about 23.5%. With the state-specific dividend yield $\widehat{q}_1 \approx 0.0310$ (about 3.1% per year), the implied total-return drift is $\widehat{\mu}_{\text{tot},1} = 0.2656$, corresponding to an expected annual total return of about 26.6%. Volatility is low in this state, $\widehat{\sigma}_1 = 0.0707$ (about 7.1% annually), with a narrow 95% confidence interval [0.0680, 0.0734]. The confidence interval for $\widehat{\mu}_{\text{cap},1}$, [0.1990, 0.2703], lies well above zero. State 1 is therefore naturally interpreted as a high-growth, low-volatility bull state.

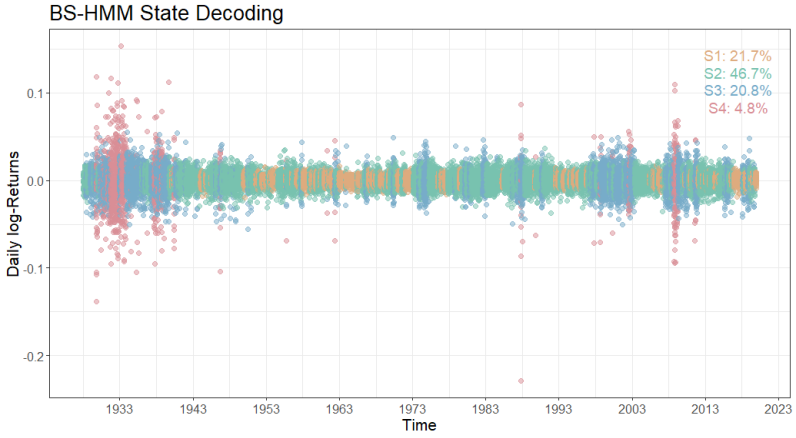
State 2 is also expansionary but more moderate. The estimated capital-gains drift $\widehat{\mu}_{\text{cap},2} = 0.1062$ corresponds to an expected annual price appreciation of about 10.6%, with a 95% confidence interval [0.0621, 0.1502] that lies strictly above zero. Together with $\widehat{q}_2 \approx 0.0348$ (about 3.5% per year), this yields $\widehat{\mu}_{\text{tot},2} = 0.1410$, or an expected annual total return of about 14.1%. Volatility is higher than in state 1 but still moderate, $\widehat{\sigma}_2 = 0.1270$ (about 12.7% annually) and below the constant-volatility BSM benchmark $\widehat{\sigma} \approx 0.1883$ (see Table 10). State 2 can be interpreted as a normal bull regime with positive returns and moderate risk.

State 3 represents a shift toward adverse market conditions. The capital-gains drift $\widehat{\mu}_{\text{cap},3} = -0.1031$ implies an expected annual price decline of about 10.3%. Even after adding $\widehat{q}_3 \approx 0.0321$ (about 3.2%), the implied total-return drift remains negative at $\widehat{\mu}_{\text{tot},3} = -0.0710$ (about -7.1% annually). The confidence interval for $\widehat{\mu}_{\text{cap},3}$, [$-0.2193$, 0.0132], narrowly straddles zero. Volatility increases substantially to $\widehat{\sigma}_3 = 0.2301$ (about 23.0% annually), which is above the BSM benchmark and indicative of elevated market turbulence. State 3 is therefore consistent with a mild bear or correction regime.

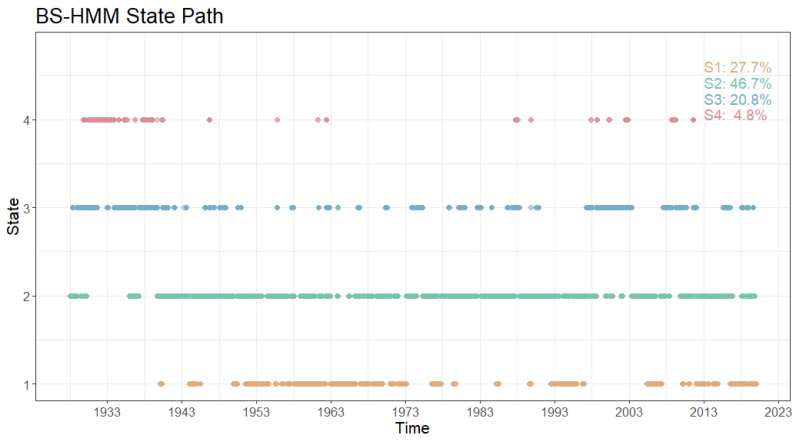
State 4 corresponds to the most adverse and volatile conditions. The capital-gains drift $\widehat{\mu}_{\text{cap},4} = -0.3818$ implies an expected annual price decline of about 38.2%. Even with the higher dividend yield $\widehat{q}_4 \approx 0.0512$ (about 5.1%), the total-return drift is highly negative at $\widehat{\mu}_{\text{tot},4} = -0.3306$ (about -33.1% annually). The confidence interval for $\widehat{\mu}_{\text{cap},4}$, [$-0.9260$, 0.1623], is wide, reflecting both the rarity and severity of such episodes. Volatility is extreme, $\widehat{\sigma}_4 = 0.5651$ (about 56.5% annually), with a confidence interval [0.5329, 0.5974]. State 4 is therefore naturally interpreted as a crisis regime.

We observe in Figure 19(a) and Figure 19(b) that the HMM successfully identifies major episodes of market turbulence, including the Wall Street Crash of 1929 and the 1937-1938 recession, the early-1980s recession, Black Monday in 1987, the dot-com episode in the late 1990s and early 2000s and the 2008 financial crisis. In these periods, the constant-parameter BSM generates particularly large residuals, whereas the BS-HMM allocates observations to high-volatility bear regimes. Episodes of the most turbulent bear market conditions (state 4) are relatively rare, whereas the milder bear regime (state 3) occurs more frequently. The Markov chain visits state 4 roughly 4.8% of the time and state 3 about 20.8%, numbers that are closely aligned with the stationary distribution $\widehat{\boldsymbol{\delta}} = (0.2766, 0.4675, 0.2080, 0.0479)$ implied by the estimated transition matrix.

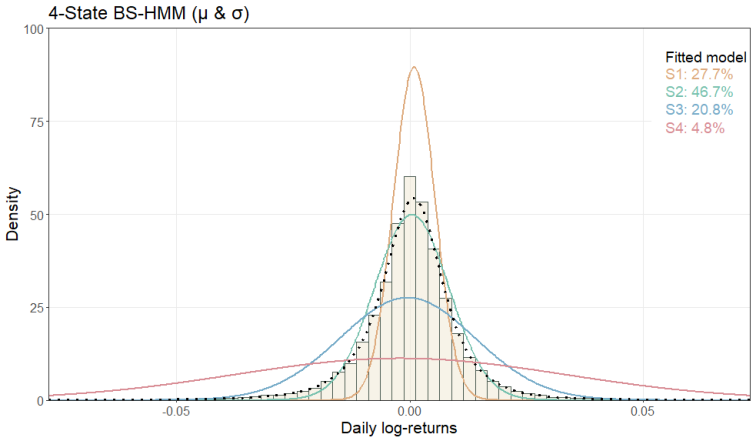
The estimated transition probability matrix $\widehat{\mathbf{\Gamma}}$ also supports an ordered regime interpretation. The diagonal entries are close to one, with $\widehat{\gamma}_{11} = 0.9649$, $\widehat{\gamma}_{22} = 0.9683$, $\widehat{\gamma}_{33} = 0.9632$ and $\widehat{\gamma}_{44} = 0.9483$. This implies expected regime durations on the order of one to one-and-a-half months in all four states, with crisis episodes somewhat shorter-lived. The off-diagonal structure is close to tridiagonal: transitions occur primarily between neighbouring regimes in terms of severity (1 to 2, 2 to 3 and 3 to 2 or 4), while direct jumps between the most extreme states are essentially absent.



((a)) Daily log-returns over time, coloured by the Viterbi most likely state sequence for the 4-state BS-HMM with $\mu_{cap,i}$ and σ_i state-dependent.



((b)) Viterbi-decoded most likely state sequence over time for the 4-state BS-HMM with $\mu_{cap,i}$ and σ_i state-dependent.



((c)) Histogram of daily log-returns and state-dependent densities for the 4-state BS-HMM with $\mu_{cap,i}$ and σ_i state-dependent.

Figure 19: *BS-HMM presentation.*

In particular, conditional on leaving state 4, the chain moves to state 3 with probability essentially equal to one. Transitions from state 4 directly to the bull states 1 or 2 are ruled out by the estimates. This implies that recoveries from crisis conditions typically proceed through an intermediate, still adverse regime rather than shifting directly to expansionary conditions. This pattern is visible in Figure 19(b), where state 4 episodes are generally preceded and followed by spells in state 3. Interestingly, the model does assign a vanishingly small value of $\gamma_{14} = 0.0007$. That is, given the model, the probability of state-transitioning from the extremely positive expansion state 1 to the extremely pessimistic state 4, is not 0.

The model also appears to identify expansion regimes in an economically sensible manner. Periods associated with the most favorable bull state 1 are prominent during well-known expansionary episodes, such as the post-war boom of the 1950s-1960s and much of the 1982-2000 secular bull market. As expected, the more moderate bull state 2 is the most frequently occupied regime. The chain spends approximately 46.7% of the sample in state 2, fairly evenly spread across the sample, as shown in Figure 19(b).

The transition probability matrix is visualized in Figure 20.

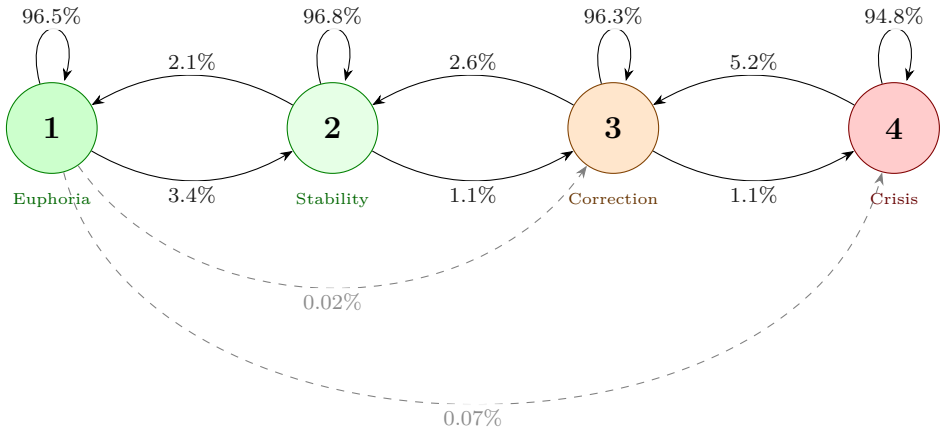


Figure 20: *Transition graph corresponding to the estimated transition matrix* $\widehat{\mathbf{\Gamma}}$*, with edges shown for* $\widehat{\Gamma}_{ij} > 0$.

Taken together, the four states provide a coherent ordering of market conditions. States 1 and 2 represent expansionary bull-market phases with positive expected returns and relatively low volatility. State 1 captures particularly strong and stable booms, while state 2 corresponds to more typical growth periods. States 3 and 4 capture contractionary bear-market phases. State 3 is a mild bear state with near-zero to moderately negative expected returns and high volatility, while state 4 corresponds to a deep crisis state with very negative expected returns and extraordinary volatility. Volatility increases monotonically as one moves from the most optimistic to the most adverse state, consistent with the empirical observation that volatility is higher in downturns. Dividend yields remain close to 3–3.5% in the first three states, but increase to roughly 5% in the crisis regime, consistent with depressed equity prices during severe downturns.

Relative to the constant-parameter BSM benchmark, which imposes a single drift and volatility $(\widehat{\mu}_{\text{cap}}, \widehat{\sigma})$ over the full sample, the BS-HMM offers a substantially richer description of the time-varying risk–return trade-off. Instead of a single average regime, the model allows the S&P 500 to switch between distinct bull and bear states with markedly different expected returns and volatility, including a tranquil high-growth regime and a turbulent crisis regime. This state dependence is empirically important. Periods of strong performance are associated with lower volatility, while severe downturns coincide with sharply higher volatility and only partial compensation through higher dividend yields. The BS-HMM therefore provides a natural benchmark for the more elaborate regime-switching specifications considered in the subsequent analysis and illustrates the limitations of the constant-parameter BSM when both expected returns and risk vary substantially over time.

Black-Scholes Continuous State-Space Model We now present the BS-SSM$_\beta$ in Equation 28.

Parameter estimates are reported in Table 12 and the remaining BS-SSM and BS-SSM$_\beta$ model parameter estimates are reported in Section A.4. The latent factor C_t is highly persistent, with $\widehat{\rho} = 0.9827$ (95% CI [0.9794, 0.9861]), which corresponds to a shock half-life of roughly $\ln 0.5 / \ln 0.9827 \approx 40$ trading days. Together with $\widehat{\sigma}_\varepsilon = 0.0769$, the state process evolves gradually and tracks medium-run variation in market conditions.

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.00170
$\widehat{\sigma}_\varepsilon$	0.0769 (0.0769, 0.0769)	0.00000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.00092
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.01753
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.00494
$\widehat{\beta}_\mu$	-0.2459 (-0.2459, -0.2459)	0.00000
$\widehat{\beta}_\sigma$	1.2824 (1.2824, 1.2824)	0.00000

Table 12: *BS-SSM$_\beta$ estimated using an $m = 100$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

As discussed in the identifiability remark for the BS-SSM$_\beta$, the parametrisation is not globally identifiable because the latent factor and its loadings are invariant under a joint scale and sign transformation. This shows up numerically as an ill-conditioned Hessian in directions associated with $(\sigma_\varepsilon, \beta_\mu, \beta_\sigma)$, i.e. the likelihood is nearly flat along a reparameterization ridge. In the present implementation, the delta-method transformation to the natural scale produces slightly negative variance estimates for β_μ and β_σ, which are truncated at zero and therefore appear as vanishing standard errors in Table 12. These zero entries do not indicate negligible estimation uncertainty; rather, they reflect that Hessian-based standard errors are not reliable for $(\sigma_\varepsilon, \beta_\mu, \beta_\sigma)$ in this specification. By contrast, the level parameters $(\rho, \mu_{\text{cap}}, q, \sigma)$ are well-behaved and can be interpreted in the usual way. Moreover, all likelihood- and distribution-based objects used below depend only on the implied drift and volatility paths (μ_t, σ_t) and are therefore invariant to the latent-factor scale and sign indeterminacy. This includes AIC/BIC, pseudo-residuals, forecast distributions and forecast-error measures.

The baseline annualized capital-gains drift at $C_t = 0$ is $\widehat{\mu}_{\text{cap}} = 0.0909$, corresponding to an expected annual price appreciation of about 9.1%. The dividend yield is estimated as $\widehat{q} \approx 0.0340$ (about 3.4% per year), implying a baseline total-return drift of $\widehat{\mu}_{\text{tot}} = 0.1249$, i.e. an expected annual total return of about 12.5% when the latent factor is neutral. The baseline volatility level is $\widehat{\sigma} = 0.1306$ (about 13.1% annually). Relative to the constant-volatility BSM benchmark, this lower level reflects that extreme movements are captured through time variation in σ_t rather than by inflating a single constant volatility parameter.

The loadings $\widehat{\beta}_\mu$ and $\widehat{\beta}_\sigma$ determine how the latent factor maps into conditional drift and volatility. Under the chosen orientation of the factor, $\widehat{\beta}_\mu < 0$ implies that higher values of C_t reduce the conditional capital-gains drift, while $\widehat{\beta}_\sigma > 0$ implies that conditional volatility increases when

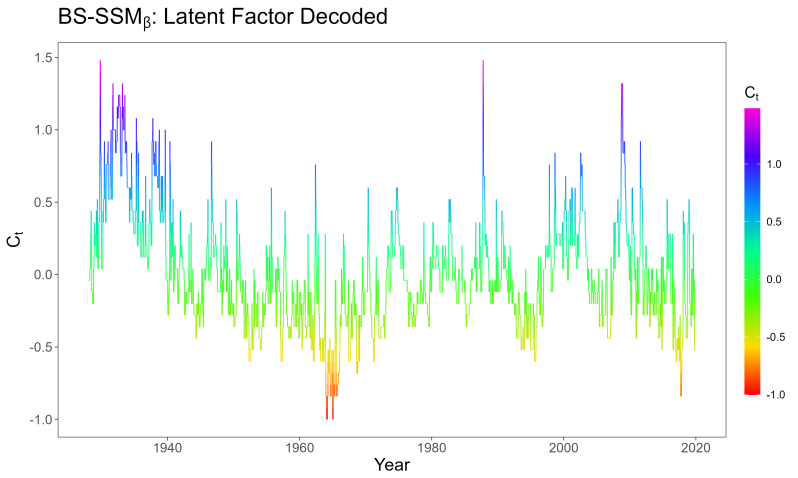
C_t is high. The opposite signs therefore mean that the latent factor acts as a "stress" index in the sense that movements in C_t are associated with a deterioration of the conditional risk-return trade-off: periods in which the model implies higher conditional volatility also tend to be periods in which the model implies lower conditional expected returns.

This co-movement is economically plausible for equity returns. In downturns and crisis episodes, volatility typically rises sharply while expected returns are depressed (or, more conservatively, conditional drift is lower) due to heightened uncertainty, tighter financial conditions and elevated risk premia. Conversely, during tranquil expansions, volatility tends to be lower and expected returns higher. The BS-SSM$_\beta$ captures this asymmetry through a single persistent factor that jointly shifts both the mean and the volatility.

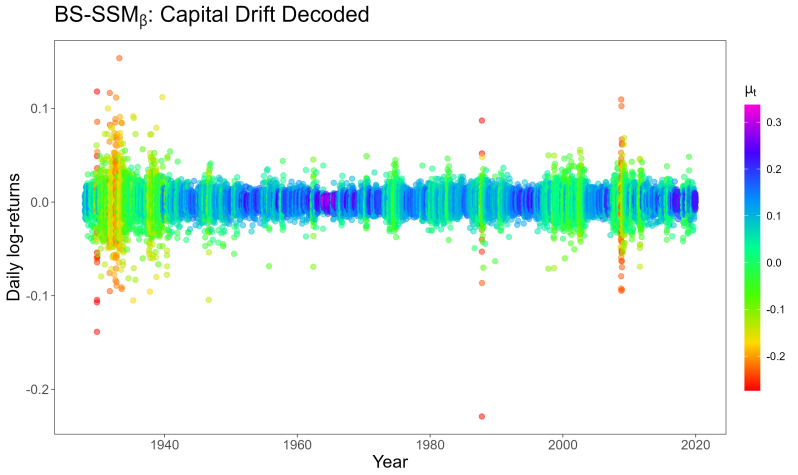
Because C_t is latent, its absolute scale is not intrinsically identified and there is also a sign invariance: mapping $C_t \mapsto -C_t$ and $(\beta_\mu, \beta_\sigma) \mapsto (-\beta_\mu, -\beta_\sigma)$ leaves (μ_t, σ_t) unchanged. For this reason, the labels "high" and "low" C_t depend on the chosen orientation. What is identified and interpretable is the relative structure that the factor moves drift and volatility in opposite directions, i.e. that the model induces a negative association between μ_t and σ_t through the common driver C_t (under the fitted orientation). The empirical content thus lies in the implied co-movement: periods with elevated C_t coincide with lower conditional drift and higher conditional volatility, while tranquil periods correspond to lower C_t, higher drift and lower volatility.

In this sense, the BS-SSM$_\beta$ provides a continuous analogue of the bull and bear regimes identified by the BS-HMM. Rather than switching discretely between a finite set of states, returns evolve along a persistent latent factor that jointly modulates drift and volatility, allowing for gradual transitions between favorable and adverse market conditions.

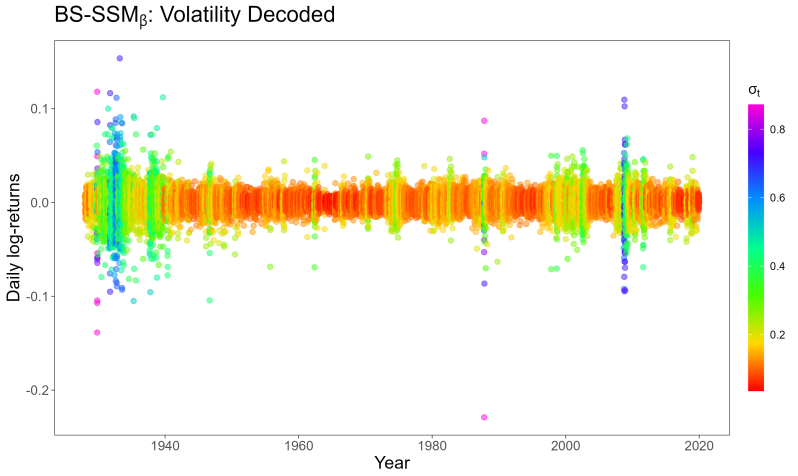
From Figure 21(a), the latent factor C_t rises sharply during major crisis episodes and is low in tranquil expansionary periods. With $\beta_\mu < 0$ and $\beta_\sigma > 0$, these movements map directly into the conditional risk–return profile: $\mu_t = \mu + \beta_\mu C_t$ decreases when C_t is high and increases when C_t is low. As shown in Figure 21(b), the most adverse configuration occurs on 28 October 1929, at the onset of the Wall Street Crash, where the model implies an annualised capital-gains drift of about $\mu_t \approx -0.27$. During the calm mid-1960s expansion, the drift peaks around $\mu_t \approx 0.34$ on 6 February 1964. Volatility moves in the opposite direction. By construction, $\sigma_t = \sigma \exp(\beta_\sigma C_t)$ is positive and increasing in C_t. In Figure 21(c), σ_t reaches roughly 0.87 on 28 October 1929 and falls to about 0.04 on 6 February 1964. Taken together, Figure 21(a) and Figure 21(c) show that the BS-SSM$_\beta$ generates crisis-like periods with high volatility and depressed expected returns and calm periods with low volatility and elevated expected returns. The continuous state-space specification therefore distils the discrete regime structure of the BS-HMM into a single latent factor that tracks the build-up and subsequent easing of market stress over time. Furthermore, unlike the BS-HMM, the BS-SSM$_\beta$ can more naturally adapt to market changes that differ from usual regimes.



((a)) Latent factor C_t over time.



((b)) Daily log-returns over time, coloured by the Viterbi-decoded most likely grid-point sequence under $\mu_t = \mu + \beta_\mu C_t$.



((c)) Daily log-returns over time, coloured by the Viterbi-decoded most likely grid-point sequence under $\sigma_t = \sigma \exp(\beta_\sigma C_t)$.

Figure 21: *BS-SSM$_\beta$ diagnostics.*

Relative to the constant-parameter BSM and the discrete-state BS-HMM, the BS-SSM$_\beta$ provides a parsimonious specification that allows both expected returns and risk to vary over time. A single persistent latent factor generates market conditions that correspond closely to the expansionary, low-volatility and crisis-like, high-volatility regimes identified by the BS-HMM, while allowing transitions between these configurations to occur gradually rather than through discrete jumps. The information criteria indicate that the BS-SSM$_\beta$ improves upon the simpler BS-SSM without factor loadings. At the same time, the model retains a transparent economic interpretation: time variation in drift and volatility is driven by a slowly evolving latent component that captures persistent changes in the return environment.

4.3 Forecast

State-space models can exhibit strong in-sample fit yet deliver noticeably weaker out-of-sample forecasting performance. This pattern is well documented in finance, where models may retrospectively rationalise regime shifts or volatility dynamics but fail to improve predictive accuracy relative to simple benchmarks on new data [17, 6]. For example, [17] find that a Markov-switching exchange-rate model fits the estimation sample well but does not outperform a random-walk benchmark in forecasting. More recent evidence similarly suggests that even flexible HMM-based specifications with multiple states, despite higher in-sample likelihood, often yield only modest improvements in out-of-sample accuracy [24]. Related discrepancies are also reported outside finance, consistent with the general tendency of complex models to capture idiosyncratic in-sample variation without improving predictive generalisation [33]. Common explanations include overfitting to transient patterns, additional uncertainty from imperfectly inferred latent states and the inherent difficulty of anticipating abrupt structural change. Accordingly, despite their sophistication and strong in-sample fit, HMMs and related state-space models often provide limited gains in short-horizon out-of-sample forecasting [6].

To assess one-step-ahead point forecasting performance for the three BS-type specifications, we report the mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE) in Table 13.

	MSE		RMSE		MAE	
Model	MSE	Rel. err. [%]	RMSE	Rel. err. [%]	MAE	Rel. err. [%]
BSM	**$10^{-4} \times 1.82$**	—	**0.0135**	—	0.00876	0.0251
BS-HMM	$10^{-4} \times 1.82$	0.162	0.0135	0.0809	0.00876	0.0154
BS-SSM$_\beta$	$10^{-4} \times 1.82$	0.120	0.0135	0.0598	**0.00876**	—

Table 13: *Out-of-sample one-step-ahead prediction errors at* **three significant digits**. *Errors are in daily log-return units. Relative error is the percentage increase over the best-performing model for each metric that is marked in bold KU-red.*

The results in Table 13 show that the three specifications perform almost indistinguishably in one-step-ahead point forecasting on the test sample. The constant-parameter BSM attains the lowest MSE and RMSE, equal to $10^{-4} \times 1.82$ and 0.0135 and therefore serves as the benchmark for the relative-error columns. The BS-HMM and BS-SSM$_\beta$ are only marginally worse in squared-error terms: their MSEs exceed the BSM by approximately 0.16% and 0.12% and their RMSEs by less than 0.1%, i.e. differences at the fourth decimal place in daily log-return units. For MAE, BS-SSM$_\beta$ achieves the smallest value (0.00876), with the BSM and BS-HMM higher by about 0.03% and 0.02%, respectively. In absolute terms, these differences correspond to changes in average forecast error on the order of 10^{-6}. Overall, neither the discrete state structure of the BS-HMM nor the continuous latent factor in the BS-SSM$_\beta$ delivers a materially better one-step-ahead point forecast than the simplest BSM benchmark, consistent with the evidence cited above for short-horizon forecasts.

Table 14 complements the point-error comparison by summarising the full one-day-ahead forecast distributions at a range of horizons and forecast origins. For the BSM, the forecast mode, median, and mean are identical across all horizons and origins. This reflects the time-homogeneous Gaussian structure of the model: conditional on the estimated parameters, the one-day-ahead return distribution is invariant to the forecast origin. The nominal 90% forecast interval is therefore constant at approximately $(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$, which provides a horizon-independent benchmark for daily return risk under the constant-parameter specification.

Year	Horizon	Mode	Median	Mean	90% interval
BSM					
2020	1 day	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
2020	1 week	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
2020	1 month	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
2020	3 months	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
2020	1 year	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
2022	3 years	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
2025	full horizon (≈ 5.66 yrs)	2.25×10^{-4}	2.25×10^{-4}	2.25×10^{-4}	$(-1.93 \times 10^{-2},\ 1.97 \times 10^{-2})$
BS-HMM					
2020	1 day	8.99×10^{-4}	8.54×10^{-4}	8.55×10^{-4}	$(-7.23 \times 10^{-3},\ 8.87 \times 10^{-3})$
2020	1 week	8.97×10^{-4}	8.46×10^{-4}	7.87×10^{-4}	$(-8.14 \times 10^{-3},\ 9.59 \times 10^{-3})$
2020	1 month	8.56×10^{-4}	7.28×10^{-4}	5.92×10^{-4}	$(-1.11 \times 10^{-2},\ 1.18 \times 10^{-2})$
2020	3 months	8.03×10^{-4}	5.87×10^{-4}	3.58×10^{-4}	$(-1.47 \times 10^{-2},\ 1.46 \times 10^{-2})$
2020	1 year	7.80×10^{-4}	5.19×10^{-4}	2.28×10^{-4}	$(-1.69 \times 10^{-2},\ 1.65 \times 10^{-2})$
2022	3 years	7.79×10^{-4}	5.18×10^{-4}	2.27×10^{-4}	$(-1.70 \times 10^{-2},\ 1.65 \times 10^{-2})$
2025	full horizon (≈ 5.66 yrs)	7.79×10^{-4}	5.18×10^{-4}	2.27×10^{-4}	$(-1.70 \times 10^{-2},\ 1.65 \times 10^{-2})$
BS-SSM$_\beta$					
2020	1 day	9.89×10^{-4}	8.91×10^{-4}	8.30×10^{-4}	$(-6.92 \times 10^{-3},\ 8.40 \times 10^{-3})$
2020	1 week	1.02×10^{-3}	8.73×10^{-4}	7.96×10^{-4}	$(-7.51 \times 10^{-3},\ 8.86 \times 10^{-3})$
2020	1 month	1.10×10^{-3}	8.21×10^{-4}	6.80×10^{-4}	$(-9.59 \times 10^{-3},\ 1.05 \times 10^{-2})$
2020	3 months	1.11×10^{-3}	6.98×10^{-4}	4.87×10^{-4}	$(-1.32 \times 10^{-2},\ 1.36 \times 10^{-2})$
2020	1 year	9.89×10^{-4}	5.49×10^{-4}	3.08×10^{-4}	$(-1.68 \times 10^{-2},\ 1.67 \times 10^{-2})$
2022	3 years	9.83×10^{-4}	5.43×10^{-4}	3.01×10^{-4}	$(-1.69 \times 10^{-2},\ 1.69 \times 10^{-2})$
2025	full horizon (≈ 5.66 yrs)	9.83×10^{-4}	5.43×10^{-4}	3.01×10^{-4}	$(-1.69 \times 10^{-2},\ 1.69 \times 10^{-2})$

Table 14: *Multi-horizon forecast summaries. Horizons are measured from the end of the estimation sample (2019-12-31). Nominal 90% forecast intervals are given in parentheses.*

By contrast, the BS-HMM and BS-SSM$_\beta$ generate horizon-dependent forecast distributions

through their latent state dynamics. At short horizons (1 day and 1 week from 2019-12-31), both models imply substantially higher conditional means and noticeably tighter 90% intervals than the BSM. For example, one day ahead the BSM forecast mean is 2.25×10^{-4}, whereas the BS-HMM and BS-SSM$_\beta$ means are 8.55×10^{-4} and 8.30×10^{-4}, respectively. At the same time, the BS-HMM 90% interval $(-7.23 \times 10^{-3}, 8.87 \times 10^{-3})$ and the BS-SSM$_\beta$ interval $(-6.92 \times 10^{-3}, 8.40 \times 10^{-3})$ are far narrower than the BSM band $(-1.93 \times 10^{-2}, 1.97 \times 10^{-2})$. This reflects that, at the end of the estimation sample, the filtered state distributions in both state-space models assign high probability to benign configurations, yielding a concentrated short-horizon predictive distribution relative to the unconditional BSM benchmark.

As the horizon increases to one month and three months, the forecast means under the BS-HMM and BS-SSM$_\beta$ decline toward smaller positive values, while the 90% intervals widen monotonically. At a one-year horizon, the BS-HMM mean has fallen to 2.28×10^{-4} and the BS-SSM$_\beta$ mean to 3.08×10^{-4}, both close to the BSM benchmark of 2.25×10^{-4}. The corresponding 90% intervals, $(-1.69 \times 10^{-2}, 1.65 \times 10^{-2})$ for the BS-HMM and $(-1.68 \times 10^{-2}, 1.67 \times 10^{-2})$ for the BS-SSM$_\beta$, are also close to the BSM band, with all three lying in the range of approximately ± 1.6–2.0% per day.

At horizons of three years and the full remaining sample (≈ 5.66 years), the BS-HMM and BS-SSM$_\beta$ rows have essentially stabilised: the forecast means and 90% intervals at three years are numerically almost identical to those at the full horizon. This indicates that the predictive densities have converged to their stationary counterparts, consistent with Equation 24, which states that the h-step-ahead forecast distribution converges to the marginal stationary density as $h \to \infty$. The table suggests that this convergence occurs at horizons on the order of one to three years, after which the incremental information from the latent state becomes negligible for a one-day-ahead forecast.

The horizons in Table 14 (one day, one week, one month, three months, one year, three years and the full out-of-sample period) span short-term trading and risk-management horizons as well as medium- and long-term investment horizons. Overall, the results suggest that while the BS-HMM and BS-SSM$_\beta$ produce richer state-dependent predictive densities and more informative short-horizon uncertainty quantification, their incremental gains in one-step-ahead point forecasting accuracy relative to the BSM are modest, consistent with the broader forecasting evidence cited above.

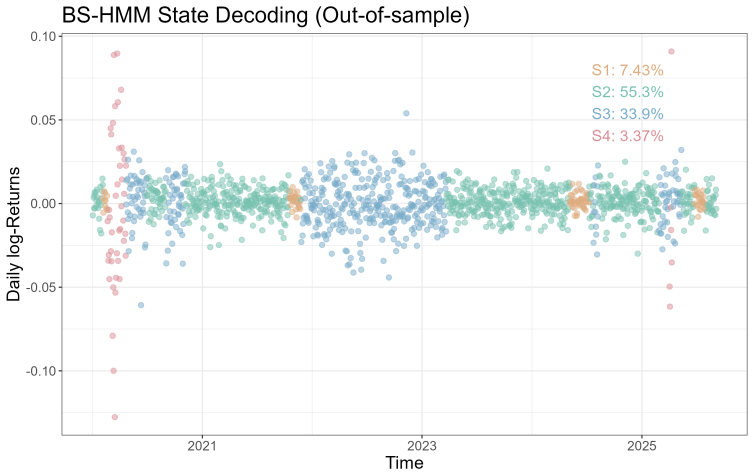
Over the next two pages we shortly visualize the Viterbi algorithm on the out-of-sample data using the models calibrated on the training data for both the BS-HMM and the BS-SSM$_\beta$.

The BS-HMM Viterbi-decoded most likely state sequence is shown in Figure 22(a) and Figure 22(b). In the out-of-sample period, state 4 occurs almost exclusively during the initial COVID-19 shock, where the model classifies the market as being in the pessimistic crisis regime. As conditions stabilize, the decoded path moves back through the correction regime (state 3) and into the stability regime (state 2), consistent with a gradual normalization.

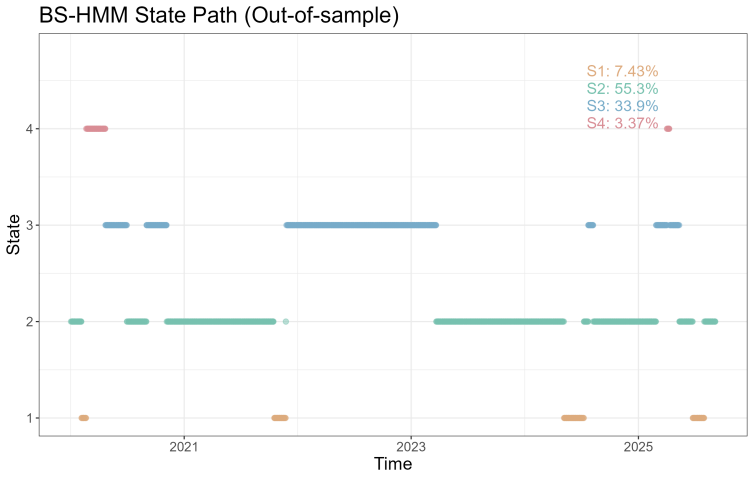
During 2022, the sequence is dominated by the correction regime, aligned with the inflation and policy backdrop; CPI inflation reached 9.1% year-on-year in June 2022 [55]. From 2023 onward, the model primarily assigns observations to the stability regime with intermittent corrections, consistent with disinflation and a strong equity rally [19]. In 2025, the decoded path exhibits frequent switching between states 2 and 3 amid heightened political and policy uncertainty; CPI inflation was 2.7% year-on-year in November 2025 [54].

Relative to the in-sample results, the out-of-sample path assigns relatively little mass to the most optimistic regime (state 1) and instead concentrates on states 2 and 3, which is consistent with a more turbulent business-cycle environment. Finally, the fitted density in Figure 22(c) appears slightly weaker than the in-sample fit in Figure 19(c).

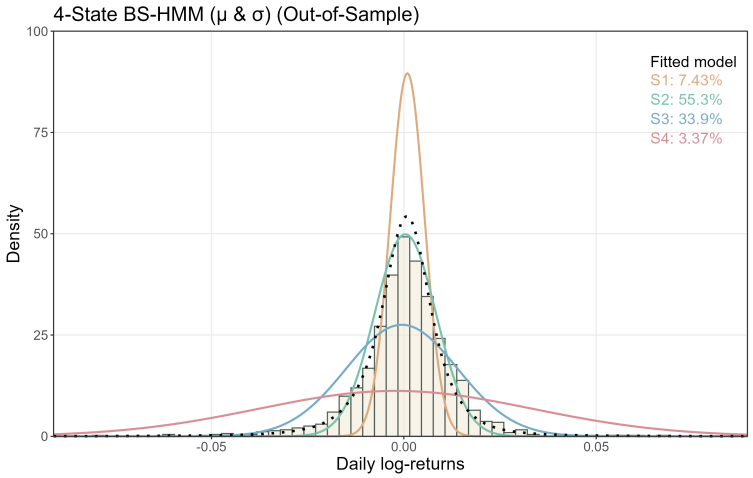
Overall, the BS-HMM provides a useful regime-based description for risk characterization relative to the BSM.



((a)) Log-returns out-of-sample over time, coloured by the Viterbi most likely state sequence for the 4-state BS-HMM with $\mu_{cap,i}$ and σ_i state-dependent.



((b)) Viterbi most likely state sequence out-of-sample over time for the 4-state BS-HMM with $\mu_{cap,i}$ and σ_i state-dependent.

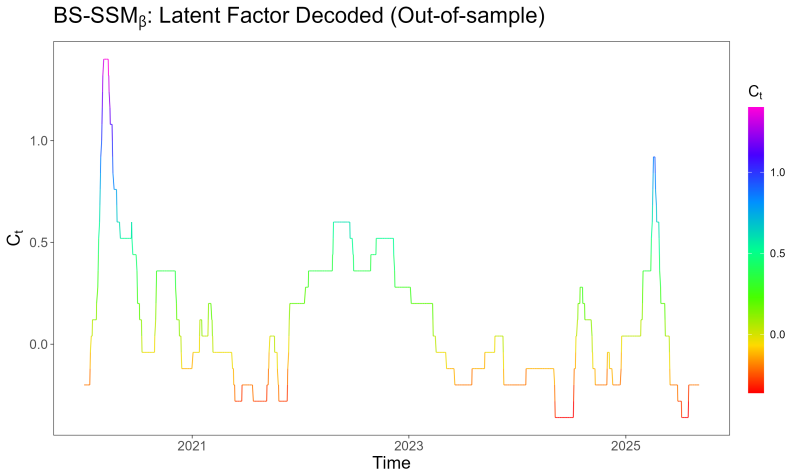


((c)) Histogram of log-returns out-of-sample and state-dependent densities for the 4-state BS-HMM with $\mu_{cap,i}$ and σ_i state-dependent.

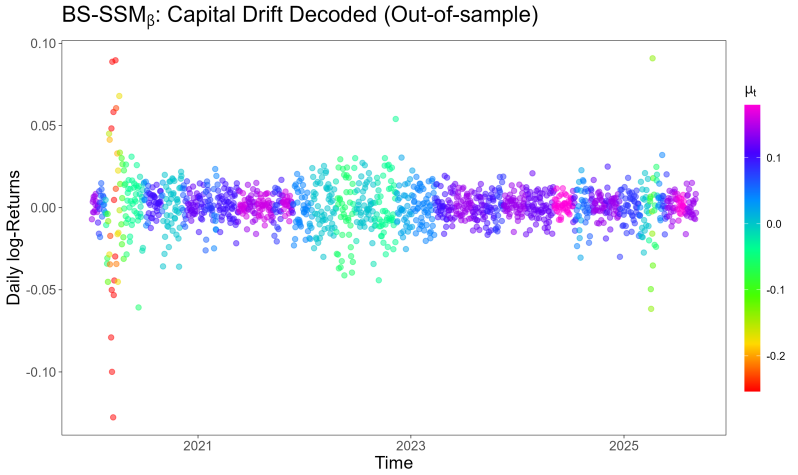
Figure 22: *BS-HMM out-of-sample presentation.*

The out-of-sample diagnostics for the BS-SSM$_\beta$ in Figure 23 are consistent with the in-sample interpretation. Elevated values of the latent factor C_t map into higher conditional volatility and lower conditional drift through $\widehat{\beta}_\sigma > 0$ and $\widehat{\beta}_\mu < 0$. In particular, σ_t spikes at the beginning of the test period, is moderately elevated toward the end of 2022 and rises again around the start of 2025, while remaining comparatively low otherwise. The conditional drift μ_t shows the corresponding pattern, taking very negative values (around -0.25) at the start of the test period, falling to moderately negative levels (around -0.10) toward the end of 2022 and declining again around the start of 2025, while being closer to neutral or positive (roughly between 0 and 0.18) in the remaining periods.

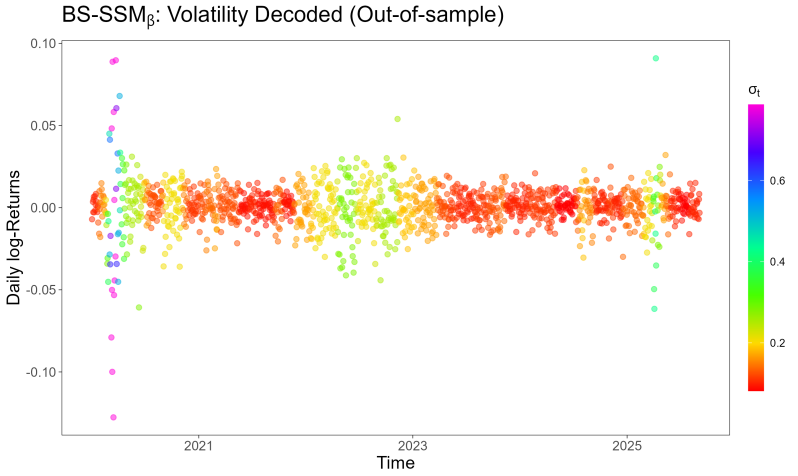
Since the latent factor is not uniquely identified in scale or sign, its absolute level is not interpreted as a cardinal quantity. However, its relative movements over time are meaningful within a fixed normalization. The elevated values of C_t at the beginning of the test period, toward the end of 2022 and around the start of 2025 align with the periods where the model implies low μ_t and high σ_t. In contrast, lower values of C_t coincide with more tranquil conditions. Overall, the BS-SSM$_\beta$ provides a coherent out-of-sample description of time variation in the risk-return trade-off through persistent movements in a single latent factor.



((a)) Latent factor C_t over time.



((b)) Daily log-returns over time, coloured by the Viterbi-decoded most likely grid-point sequence under $\mu_t = \mu + \beta_\mu C_t$.



((c)) Daily log-returns over time, coloured by the Viterbi-decoded most likely grid-point sequence under $\sigma_t = \sigma \exp(\beta_\sigma C_t)$.

Figure 23: *BS-SSM$_\beta$ out of sample presentation.*

5 Discussion

The empirical investigation conducted in this thesis was motivated by a fundamental dichotomy in financial econometrics, that is, the tension between model parsimony and descriptive realism. The Black-Scholes model (BSM), while serving as the bedrock of modern derivatives pricing through its elegant closed-form solutions and assumption of geometric Brownian motion, essentially posits a market environment characterized by constant probabilistic laws. However, the historical record of the S&P 500 index data at hand, spanning nearly a century from 1927 to 2025, reveals a data generating process defined not by constancy, but by profound structural instability. This thesis has sought to bridge the gap between the tractable but rigid BSM and the complex reality of financial markets by introducing regime-switching dynamics through both discrete Hidden Markov Models (BS-HMM) and continuous State-Space Models (BS-SSM). The results obtained offer a nuanced perspective on the utility of these extensions, revealing a striking divergence between their in-sample explanatory power and their out-of-sample predictive efficacy.

The initial examination of the standard Black-Scholes model provided a necessary baseline, confirming its limitations in capturing the distributional properties of long-horizon asset returns. The analysis of residuals derived from the constant-parameter BSM demonstrated unambiguous evidence of leptokurtosis and volatility clustering. These are stylized facts that are well-documented in the literature but critically undermine the model's assumption of independent and identically distributed Gaussian log-returns. The prevalence of extreme outliers in the standardized residuals, particularly during historically significant periods such as the Wall Street Crash of 1929, the Black Monday crash of 1987 and the Global Financial Crisis of 2008, underscores the model's inability to adapt to changing risk environments. Under the BSM framework, these events are treated as statistically impossible anomalies rather than integral components of the market's stochastic process. The persistence of these large residuals suggests that a single volatility parameter $\hat{\sigma} \approx 18.83\%$ acts as an indiscriminate average, overestimating risk during tranquil periods and catastrophically underestimating it during crises. Consequently, the assumption of a constant capital-gains drift $\hat{\mu}_{cap}$ and volatility $\hat{\sigma}$ imposes a rigidity that masks the distinct economic states driving asset prices.

The transition to regime-switching frameworks was predicated on the hypothesis that relaxing these constant parameter assumptions would resolve the BSM's distributional inadequacies. The implementation of the discrete Black-Scholes Hidden Markov Model (BS-HMM) yielded arguably the most economically interpretable results of this study. By allowing the drift and volatility parameters to assume state-dependent values governed by a latent Markov chain, the model successfully decomposed the aggregate return series into distinct market phases. The superior fit of the 4-state BS-HMM, selected via a combination of Information Criteria (AIC/BIC) and domain expertise, validates the existence of discrete regimes rather than a continuum of random variation. The identified states, characterized in this analysis as "Euphoria" (state 1), "Stability" (state 2), "Correction" (state 3) and "Crisis" (state 4), provide a granular mapping of market sentiment

that aligns with macroeconomic history.

A critical finding from the BS-HMM estimation is the distinct separation of volatility levels across states, ranging from a remarkably low $\hat{\sigma}_1 \approx 7.1\%$ in the euphoric state to an extreme $\hat{\sigma}_4 \approx 56.5\%$ in the crisis state. This monotonic increase in volatility as market conditions deteriorate supports the "leverage effect" hypothesis, where negative returns are correlated with increases in volatility. However, the BS-HMM refines this by showing that volatility does not merely fluctuate continuously but jumps between stable plateaus. The "Crisis" state, while rare (visited approximately 4.8% of the time), is statistically distinct from the "Correction" state (visited 20.8% of the time). This distinction is vital. Distinct from a simple high-variance state, the crisis regime is accompanied by a severe negative drift of $\hat{\mu}_{cap,4} \approx -38.2\%$. This coupling of negative drift and high volatility in the hidden states captures the "panic" behavior of markets where risk premiums explode and prices collapse simultaneously. The estimated transition probability matrix $\hat{\Gamma}$ further illuminates the structural dynamics of these shifts. The high diagonal elements indicate regime persistence, that is, markets do not flicker randomly between crash and boom but settle into states for meaningful durations (roughly one to one-and-a-half months). Furthermore, the tridiagonal structure of $\hat{\Gamma}$ suggests a "gradual adjustment" mechanism in the economy; direct transitions from the deepest crisis (state 4) to the highest euphoria (state 1) are effectively zero. The market must heal through intermediate states of correction and stability. This finding challenges jump-diffusion models that model crashes as instantaneous Poisson discontinuities; instead, our results suggest crashes are persistent regimes of high variance and negative drift.

It is pertinent to discuss the implications of model specification within the HMM framework, specifically the failure of models that allowed for state-dependent drift (μ) but imposed a common volatility (σ). The analysis revealed that such specifications performed poorly in terms of information criteria and produced residuals that failed to rectify the non-normality issues of the BSM. Mathematically, this is a consequence of the likelihood function's sensitivity to the variance term. In the Gaussian likelihood, the variance appears both in the normalization constant and the exponent. A regime-switching mean with constant variance allows the distribution to shift laterally, but it cannot account for the varying width of the return distribution. Since financial data is characterized primarily by heteroskedasticity (changing variance) rather than just changing means, a model that constrains σ to be constant across states forces the drift parameter to absorb variation it cannot explain, often resulting in erratic and economically implausible drift estimates. This confirms that volatility regimes are the primary driver of the enhanced likelihood in switching models and any successful extension of the BSM must primarily address the time-varying nature of σ.

The continuous state-space extensions offered a different theoretical advantage. Namely, the ability to model the latent market environment as a fluid, evolving factor rather than a set of discrete jumps. The BS-SSM$_\beta$ specification, which utilized a latent AR(1) process C_t to linearly load the drift and exponentially load the volatility, provided a parsimonious alternative to the

HMM. The estimated high persistence of the latent factor ($\hat{\rho} \approx 0.98$) aligns with the volatility clustering observed in the raw data. The factor loadings ($\hat{\beta}_\mu < 0$ and $\hat{\beta}_\sigma > 0$) confirm the inverse relationship between the market's "stress" level and its performance. As the latent factor C_t rises, volatility expands exponentially while expected returns contract. This continuous formulation effectively distills the discrete regimes of the HMM into a single trajectory. The visualization of the decoded latent factor C_t tracks historical crises with remarkable precision, peaking during the 1929 crash and the 2008 financial crisis and troughing during the mid-century bull markets.

However, the pure BS-SSM (without factor loading constants) exhibited significant deficiencies, notably predicting systematically high returns and failing to capture the left tail of the distribution adequately. The constraints of the pure model, where the latent state affects μ and σ symmetrically without scaling parameters, proved too restrictive. The pure BS-SSM assumes the latent state C_t impacts the drift via μe^{C_t}. Since e^{C_t} is strictly positive, this specification forces the drift to have the same sign as the base parameter μ. If μ is positive (which it generally is for equities), the model cannot generate negative expected returns, only varying magnitudes of positive returns. This structural limitation explains why the pure BS-SSM failed to capture the left tail and crisis dynamics effectively. It structurally prohibits a negative drift regime. The BS-SSM$_\beta$, by using an additive loading $\mu + \beta_\mu C_t$, allows the total drift to turn negative when C_t is sufficiently large (given $\hat{\beta}_\mu < 0$). This flexibility is mathematically necessary to model equity markets, which experience genuine contractions, not just slower growth.

Despite the superior in-sample fit and rich descriptive capabilities of both the 4-state BS-HMM and the BS-SSM$_\beta$, the out-of-sample forecasting exercise revealed a sobering paradox common in financial econometrics. The forecasting evaluation on the hold-out period (2020-2025) indicated that the regime-switching extensions offered negligible gains in point forecasting accuracy (MSE and RMSE) relative to the static BSM. In some horizons, the complex models performed marginally worse. This phenomenon warrants a deep dissection, as it touches upon the fundamental difference between explaining history and predicting the future. The primary driver of this forecasting paradox is parameter uncertainty and the challenge of latent state inference. In the BSM, the parameters μ and σ are static. While they may be biased averages, they are estimated with high precision due to the large sample size. In contrast, the BS-HMM and BS-SSM rely on inferring the current value of a hidden state based on noisy return data. This inference is probabilistic. When forecasting X_{T+1}, the model must integrate over the uncertainty of the current state C_T and its transition to C_{T+1}. If the filtering probability $\mathbb{P}\left(C_T = i | \mathbf{X}^{(T)} = \mathbf{x}^{(T)}\right)$ is imperfect, which is inevitable given the noise in daily returns, the forecast is contaminated. During transitions, specifically, the model inevitably lags. It requires a sequence of extreme returns to confirm a regime switch with high probability. By the time the model is confident the market is in a "Crisis" state, the most volatile observations may have already occurred. Consequently, for one-step-ahead point forecasts, the average prediction of the BSM is a robust, low-variance estimator that is hard to beat, even if it is theoretically misspecified.

Furthermore, the theoretical convergence of the forecast distribution explains the diminishing utility of regime-switching for long-horizon prediction. As demonstrated in the results, as the forecast horizon h increases, the predictive density $f_{X_{T+h}|X^{(T)}}$ converges to the stationary distribution of the model. This means that the information contained in the current state decays exponentially. For the BS-HMM, the mixing weights for the future forecast densities approach the stationary distribution. Therefore, at horizons beyond a few months (or years, depending on persistence), the conditional forecast becomes indistinguishable from the unconditional distribution. This explains why, in the reported forecast tables, the mean and interval estimates for the BSM, BS-HMM and BS-SSM converge at the 3-year and 5-year horizons. The memory of the regime-switching models is finite.

An important econometric issue concerns identifiability in the BS-SSM$_\beta$. The parameters $(\sigma_\varepsilon, \beta_\mu, \beta_\sigma)$ are not globally identified: rescalings of the latent factor produce the same likelihood and the same induced distribution for returns. This does not affect the fitted paths (μ_t, σ_t) or forecast distributions, but it limits structural interpretation of the individual coefficients. Accordingly, C_t is interpreted as a relative (ordinal) index of market stress rather than a cardinal factor with intrinsic units. In the implementation, this manifests as an ill-conditioned Hessian in directions associated with $(\sigma_\varepsilon, \beta_\mu, \beta_\sigma)$, leading to unstable delta-method variances and, in the tables, vanishing standard errors or truncated negative variance estimates for the loadings. If structural inference were the primary objective, an explicit normalization (e.g. fixing $\sigma_\varepsilon = 1$) would stabilise the estimates; here, we focus on likelihood-based fit and forecasting quantities that are invariant to this non-identifiability.

The issue of microstructure noise and the theoretical distinction between jumps and regimes also deserves careful consideration in light of the findings. The financial econometrics literature distinguishes between continuous diffusions and jump processes using high-frequency data. Our analysis, relying on daily closing prices, aggregates this intra-day activity. What appears as a jump in high-frequency data is modeled here as a high-volatility regime. This is a deliberate methodological choice. By modeling these events as regimes (state 4) rather than isolated jumps, we posit that the underlying economic conditions persist. A jump implies a momentary dislocation, often reversed or independent of the subsequent step. A regime implies a structural shift in risk that dictates the distribution of the next return. The persistence of the estimated crisis states supports the regime interpretation for daily data—crashes like 1929 or 2008 were not single-day outliers but sustained periods of market dysfunction. However, the presence of microstructure noise in the underlying price formation process could ostensibly mimic regime shifts. Volatility clustering at the micro-scale aggregates to the heavy tails observed in daily log-returns. While the BS-HMM captures the clustering, it does not explicitly model the noise. The finding referenced in the introduction that jumps are often erroneously identified and are rarer than commonly assumed supports our choice to use a diffusion-based switching model rather than a pure jump-diffusion. The BS-HMM suggests that what looks like a jump is often just a draw from the tail of a high-

volatility Gaussian regime (state 4), rather than a discontinuity in the price path itself.

Numerical implementation presented its own set of challenges, particularly the risk of underflow in the likelihood calculation for the continuous state-space models. The use of the normalized forward algorithm was essential to maintain numerical stability. Furthermore, the discretization of the continuous state space into m intervals introduces an approximation error. The robustness checks performed with varying m (up to 200) and boundary b_{max} suggest that our chosen grid ($m = 100$) was sufficient to stabilize the likelihood. However, the discretization essentially approximates the continuous process with a very large discrete HMM. While computationally feasible for a univariate state, this approach would suffer from the curse of dimensionality if the model were extended to include multiple latent factors (e.g., stochastic volatility and stochastic mean reverting level separately). The choice of simple quadrature for the discretization was driven by computational efficiency, but for more gradual regime switches, perhaps driven by slower, moving macroeconomic variables like interest rates or inflation, a more sophisticated integration scheme or a particle filter might offer superior resolution for the latent state trajectory.

A specific data nuance involves the dividend yield estimation. The construction of the pre-1988 dividend series using Shiller's monthly data introduces a "step function" artifact when backfilled to daily frequency. This implies that the variance of the dividend yield is artificially suppressed in the earlier part of the sample compared to the post-1988 TR-PR differencing period. While we mitigated this by estimating q as a constant (or state-dependent constant) mean, the underlying heteroskedasticity of the dividend process itself is not fully captured. Given that dividends are a relatively small component of daily price variance compared to capital gains, this approximation is likely second-order, but it remains a limitation. The state-dependent dividend yield estimates ($\hat{q}_i$) showing higher yields in crisis states ($\hat{q}_4 \approx 5.1\%$) is consistent with the mechanical effect of falling prices driving up yields (the denominator effect). This reinforcing relationship adds economic coherence to the identified states. Specifically, state 4 is not just defined by price volatility, but by the fundamental distress signal of elevated yields relative to prices.

The pseudo-residuals analysis served as the primary diagnostic tool for distributional fit. The transformation of observations to the standard normal scale $z_t = \Phi^{-1}(F(x_t))$ allows for a rigorous check of the model's assumptions. The standard BSM residuals were unequivocally non-Gaussian, failing the conceptual checks via heavy tails. The BS-HMM residuals, while an improvement, still exhibited some outliers. This suggests that even within a regime, returns may not be perfectly Gaussian, perhaps requiring a t-distribution or jump component within the states. However, the dramatic reduction in the magnitude of outliers in the switching models confirms that the majority of the fat tails in the unconditional distribution are indeed caused by the mixing of Gaussian distributions with different variances. The remaining outliers in the regime-switching pseudo-residuals likely point to idiosyncratic shocks, that is, true jumps caused by exogenous news events (e.g., geopolitical shocks, pandemics) that are instantaneous and not regime-driven. The large residuals in the pure BS-SSM, particularly the systematic overprediction of returns, provided the

clearest evidence of that model's misspecification, confirming visually what the likelihood metrics suggested numerically.

Although the MSE metrics show no advantage, the density forecasts provided by the BS-HMM and BS-SSM are fundamentally richer. The BSM assumes a constant width for the confidence intervals regardless of the current environment. The regime-switching models, conversely, provide horizon-dependent densities that expand and contract. In a high-volatility state, the 90% forecast interval is significantly wider, correctly signaling the increased value at risk. For an options trader or risk manager, knowing that the probability of staying in a crisis-regime is 95% (via $\hat{\Gamma}$) is far more valuable than a vanishing more accurate point forecast of the mean return. The static BSM would massively underprice out-of-the-money puts during a crisis because it reverts to the long-run average volatility. The BS-HMM, by conditioning on State 4, would price these options closer to their realized risk.

Therefore, the discussion must conclude that the failure of these models to improve point forecasting is not a failure of the models themselves, but a reflection of the inherent unpredictability of the conditional mean of asset returns. The signal-to-noise ratio in daily returns is notoriously low. The sophisticated machinery of HMMs and SSMs excels at characterizing the volatility and the skewness/kurtosis, which are persistent and forecastable. They do not, however, uncover a hidden deterministic trend in the drift that allows for easy profit. The findings of this thesis align with the Efficient Market Hypothesis in the weak form regarding returns, but strongly reject it regarding risk. Risk is predictable, state-dependent and clustered.

In summary, the transition from the Black-Scholes model to Markov-switching extensions represents a trade-off. We sacrifice parsimony and invite parameter uncertainty and computational complexity. In return, we gain a statistically superior description of history that captures the multi-modal, fat-tailed nature of returns. The Viterbi decoding provides a rigorous, quantitative narrative of market history that aligns with economic intuition. The 4-state BS-HMM and the factor-loaded BS-SSM$_\beta$ emerged as the preferred specifications, balancing flexibility with interpretability. While they do not function as a crystal ball for day-to-day price movements, they serve as a superior "barometer" for the prevailing risk environment. Future research could explore the integration of these regime-switching volatility forecasts into option pricing formulas directly, replacing the constant σ in the Black-Scholes formula with the regime, conditional volatility or a weighted mixture, to test if the economic value of these models is realized in the pricing of tail risk rather than the forecasting of the underlying asset.

6 Conclusion

The empirical results show that the Black-Scholes model (BSM), while theoretically coherent and computationally convenient, provides an inadequate description of S&P 500 log-returns over the long sample. In particular, the assumption of constant drift and volatility is not merely a simplifying restriction but a misspecification that obscures substantial time variation in the risk-return environment.

In-sample evidence indicates that equity returns are better characterised by persistent shifts between distinct market conditions. The estimated 4-state Black-Scholes hidden Markov model (BS-HMM) captures this feature and associates economically interpretable parameters with each regime. Two regimes correspond to expansionary conditions. In state 1, the estimated annualised capital-gains drift is approximately 23.5%, paired with a low volatility of about 7.1%. State 2 represents a more typical growth regime, with capital-gains drift around 10.6% and volatility around 12.7%.

The remaining regimes correspond to contractionary periods and, importantly, differ materially from the Gaussian tail events implied by the BSM. The state decoding in Figure 19(a) isolates major episodes of market stress, including the 1929 Crash, the 1937-1938 recession, Black Monday (1987) and the 2008 Financial Crisis, into a crisis regime (state 4) characterized by extreme volatility exceeding 56.5% and strongly negative capital-gains drift of roughly -38.2%. This regime is rare, with an empirical occupancy of approximately 4.8%. A milder bear regime (state 3), with capital-gains drift around -10.3% and volatility around 23.0%, occurs more frequently, about 20.8% of the time. These frequencies are close to the estimated stationary distribution $\widehat{\boldsymbol{\delta}} = (0.2766, 0.4675, 0.2080, 0.0479)$, reinforcing that adverse market phases are statistically distinct and persistent, even if they occupy a minority of the sample.

The transition probability matrix $\widehat{\mathbf{\Gamma}}$ further clarifies the dynamics of regime changes. The diagonal elements are close to one ($\widehat{\gamma}_{11} \approx 0.96, \ldots, \widehat{\gamma}_{44} \approx 0.95$), implying expected regime durations on the order of one to one-and-a-half months. Off-diagonal mass is concentrated on neighbouring transitions, consistent with an approximately tridiagonal structure in which the market typically moves between adjacent regimes (for example, from state 2 to state 3) rather than transitioning directly between the most benign and most adverse conditions. This discrete view is consistent with the continuous state-space specification BS-SSM$_\beta$, which identifies a highly persistent latent stress factor ($\widehat{\rho} \approx 0.98$) that shifts drift downward and volatility upward as stress increases.

Out-of-sample results provide an important qualification. Superior in-sample fit does not translate into material improvements in one-step-ahead point forecasting accuracy. In the present application, the BS-HMM and BS-SSM$_\beta$ yield point forecast errors that are essentially indistinguishable from those of the BSM at daily frequency. This outcome is consistent with existing evidence that richer latent-state structures often add parameter and state uncertainty without improving short-horizon forecasts of noisy returns.

The principal value of the regime-switching extensions therefore lies in their distributional and risk-management implications rather than in short-horizon directional prediction. The BS-HMM and BS-SSM$_\beta$ produce state-dependent and horizon-dependent forecast distributions that support risk assessment under specific adverse regimes, rather than relying on a single unconditional volatility estimate. This enables stress testing against crisis-like dynamics and provides a more informative representation of tail risk than the constant-parameter BSM.

Overall, the findings support the use of regime-switching models as tools for risk management and applications where the conditional return distribution matters, such as scenario analysis and derivative valuation, while highlighting their limited incremental benefit for one-day-ahead point forecasting of index returns. The key contribution of the framework is to quantify and distinguish between qualitatively different adverse conditions, notably between a moderate correction regime and a severe crisis regime and to track how the probability of such conditions evolves over time.

Bibliography

REFERENCES

[1] Yacine Aït-Sahalia and Jean Jacod. "Testing for Jumps in a Discretely Observed Process." In: *Annals of Statistics* 37.1 (2009), pp. 184–222.

[2] Theodore W. Anderson. *The Statistical Analysis of Time Series*. New York: John Wiley & Sons, 1971.

[3] Ole E. Barndorff-Nielsen and Neil Shephard. "Power and Bipower Variation with Stochastic Volatility and Jumps." In: *Journal of Financial Econometrics* 2.1 (2004), pp. 1–37.

[4] Tomas Björk. *Arbitrage theory in continuous time*. 4th ed. Oxford university press, 2020.

[5] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities." In: *Journal of political economy* 81.3 (1973), pp. 637–654.

[6] Tom Boot and Andreas Pick. "Optimal forecasts from Markov switching models." In: *Journal of Business & Economic Statistics* 36.4 (2018), pp. 628–642.

[7] Mark Broadie and Özgür Kaya. "Exact simulation of stochastic volatility and other affine jump diffusion processes." In: *Operations research* 54.2 (2006), pp. 217–231.

[8] Kenneth P Burnham and David R Anderson. "Multimodel inference: understanding AIC and BIC in model selection." In: *Sociological methods & research* 33.2 (2004), pp. 261–304.

[9] Kenneth P Burnham, David R Anderson, and Kathryn P Huyvaert. "AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons." In: *Behavioral ecology and sociobiology* 65 (2011), pp. 23–35.

[10] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in hidden Markov models*. Springer, 2005.

[11] Kim Christensen, Roel CA Oomen, and Mark Podolskij. "Fact or friction: Jumps at ultra high frequency." In: *Journal of Financial Economics* 114.3 (2014), pp. 576–599.

[12] D. R. Cox and E. J. Snell. "A General Definition of Residuals (with discussion)." In: *Journal of the Royal Statistical Society: Series B (Methodological)* 30 (1968), pp. 248–275.

[13] J. E. Dennis Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[14] David A Dickey and Wayne A Fuller. "Distribution of the estimators for autoregressive time series with a unit root." In: *Journal of the American statistical association* 74.366a (1979), pp. 427–431.

[15] Peter K. Dunn and Gordon K. Smyth. "Randomized Quantile Residuals." In: *Journal of Computational and Graphical Statistics* 5 (1996), pp. 236–244.

[16] Sean R Eddy. *Biological sequence analysis Probabilistic models of proteins and nucleic acids*. 1998.

[17] Charles Engel. "Can the Markov Switching Model Forecast Exchange Rates?" In: *Journal of International Economics* 36.1 (1994), pp. 151–165.

[18] Lorella Fatone et al. "The Use of Statistical Tests to Calibrate the Black-Scholes Asset Dynamics Model Applied to Pricing Options with Uncertain Volatility." In: *Journal of Probability and Statistics* 2012 (2012), Article ID 931609, 20 pages. DOI: 10.1155/2012/931609.

[19] Federal Reserve Bank of Minneapolis. *Consumer Price Index, 1913-. Historical data from the era of the modern U.S. consumer price index (CPI)*. No publication date listed on page; accessed 2025-12-22. Federal Reserve Bank of Minneapolis. URL: https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913- (visited on 12/22/2025).

[20] Federal Reserve Bank of St. Louis. *Review (Federal Reserve Bank of St. Louis)*. https://fraser.stlouisfed.org/title/820. Publication series; issues from the 2000s accessed via FRASER. 1917-2025. (Visited on 12/04/2025).

[21] William Feller. *An introduction to probability theory and its applications, Volume 2*. Vol. 2. John Wiley & Sons, 1991.

[22] Jr. Forney G. David. "The Viterbi Algorithm." In: *Proceedings of the IEEE* 61.3 (1973).

[23] Sylvia Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. New York: Springer, 2006.

[24] Arun Gopalakrishnan and Eric Bradlow. *Hidden Markov Model Backcasting Versus Forecasting Performance*. Tech. rep. Working paper, available at SSRN: https://ssrn.com/abstract=5623200. The Wharton School, 2025.

[25] James D. Hamilton. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.

[26] Zoé van Havre et al. "Overfitting hidden Markov models with an unknown number of states." In: *arXiv preprint arXiv:1602.02466* (2016).

[27] Søren Tolver Jensen and Anders Rahbek. "On the law of large numbers for (geometrically) ergodic Markov chains." In: *Econometric Theory* 23.4 (2007), pp. 761–766.

[28] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Vol. 23. Applications of Mathematics. New York, NY: Springer, 1992.

[29] Roland Langrock, Iain L MacDonald, and Walter Zucchini. "Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models." In: *Journal of Empirical Finance* 19.1 (2012), pp. 147–161.

[30] Brian G Leroux and Martin L Puterman. "Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models." In: *Biometrics* (1992), pp. 545–558.

[31] H Linhart and W Zucchini. "Model Selection, Wiley." In: *New York* (1986).

[32] Roger Lord, Remmert Koekkoek, and Dick Van Dijk. "A comparison of biased simulation schemes for stochastic volatility models." In: *Quantitative Finance* 10.2 (2010), pp. 177–194.

[33] Spyros Makridakis and Michèle Hibon. "The M3-Competition: Results, Conclusions and Implications." In: *International Journal of Forecasting* 16.4 (2000), pp. 451–476.

[34] Brett T. McClintock and Théo Michelot. "momentuHMM: R package for generalized hidden Markov models of animal movement." In: *Methods in Ecology and Evolution* 9.6 (2018), pp. 1518–1530. DOI: 10.1111/2041-210X.12995. URL: https://doi.org/10.1111/2041-210X.12995.

[35] Geoffrey J McLachlan and David Peel. *Finite mixture models*. John Wiley & Sons, 2000.

[36] Robert C Merton. "An intertemporal capital asset pricing model." In: *Econometrica: Journal of the Econometric Society* (1973), pp. 867–887.

[37] Robert C. Merton. "Option Pricing When Underlying Stock Returns Are Discontinuous." In: *Journal of Financial Economics* 3.1-2 (1976), pp. 125–144. DOI: 10.1016/0304-405X(76)90022-2.

[38] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[39] Théo Michelot, Roland Langrock, and Toby Patterson. "moveHMM: An R package for the analysis of animal movement data." In: *Computer software* (2019).

[40] John F Monahan. *Numerical methods of statistics*. Cambridge University Press, 2011.

[41] Whitney K. Newey and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." In: *Econometrica* 55.3 (1987), pp. 703–708. DOI: 10.2307/1913610.

[42] Manh Cuong Ngô, Mads Peter Heide-Jørgensen, and Susanne Ditlevsen. "Understanding narwhal diving behaviour using Hidden Markov Models with dependent state distributions and long range dependence." In: *PLoS computational biology* 15.3 (2019), e1006425.

[43] Jennifer Pohle et al. "Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement." In: *Journal of Agricultural, Biological and Environmental Statistics* 22 (2017), pp. 270–293.

[44] Anders Rahbek and Rasmus Søndergaard Pedersen. *Lecture Notes for NMAK24011U Financial Econometric Time Series Analysis (FinMetrics)*. Course lecture notes, Department of Mathematical Sciences. Aug. 2024.

[45] Cyrus A. Ramezani and Yong Zeng. "Maximum Likelihood Estimation of the Double Exponential Jump-Diffusion Process." In: *Annals of Finance* 3.4 (2007), pp. 487–507.

[46] F. W. Scholz. "Maximum likelihood estimation." In: *Encyclopedia of Statistical Sciences*. Ed. by Samuel Kotz et al. 2nd. Hoboken, NJ: Wiley, 2006, pp. 4629–4639.

[47] Robert J. Shiller. *Data Appendix: U.S. Stock Market (notes and documentation)*. Documentation of sources and splicing for price, dividends, and earnings. 2025. URL: https://www.econ.yale.edu/~shiller/data/chapt26.html (visited on 11/09/2025).

[48] Robert J. Shiller. *Online Data: U.S. Stock Market Prices, Dividends, Earnings, and CPI*. Monthly S&P price & dividend series starting in 1871. 2025. URL: https://www.econ.yale.edu/~shiller/data.htm (visited on 11/09/2025).

[49] S&P Dow Jones Indices. *FAQ: S&P 500 Dividend Points Index*. Explains the Dividend Points index and annual reset. 2023. URL: https://www.spglobal.com/spdji/en/documents/additional-material/faq-sp-500-dividend-points-index.pdf (visited on 11/09/2025).

[50] S&P Dow Jones Indices. *Index Mathematics Methodology*. URL: https://www.spglobal.com/spdji/en/methodology/article/index-mathematics-methodology/ (visited on 11/09/2025).

[51] S&P Dow Jones Indices. *Index Mathematics Methodology*. 2025. URL: https://www.spglobal.com/spdji/zh/documents/methodologies/methodology-index-math.pdf (visited on 11/09/2025).

[52] S&P Dow Jones Indices. *S&P U.S. Indices Methodology*. 2025. URL: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf (visited on 11/09/2025).

[53] Dag Tjøstheim. "Non-linear time series and Markov chains." In: *Advances in applied probability* 22.3 (1990), pp. 587–611.

[54] U.S. Bureau of Labor Statistics. *Consumer Price Index – November 2025*. U.S. Department of Labor. Dec. 2025. URL: https://www.bls.gov/news.release/cpi.htm (visited on 12/22/2025).

[55] U.S. Bureau of Labor Statistics. *Consumer prices up 9.1 percent over the year ended June 2022, largest increase in 40 years*. U.S. Department of Labor. July 2022. URL: https://www.bls.gov/opub/ted/2022/consumer-prices-up-9-1-percent-over-the-year-ended-june-2022-largest-increase-in-40-years.htm (visited on 12/22/2025).

[56] Ingmar Visser and Maarten Speekenbrink. "depmixS4: An R Package for Hidden Markov Models." In: *Journal of Statistical Software* 36.7 (2010), pp. 1–21. DOI: 10.18637/jss.v036.i07. URL: https://www.jstatsoft.org/v36/i07/.

[57] Andrew J. Viterbi. "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm." In: *IEEE Transactions on Information Theory* 13.2 (1967).

[58] Larry Wasserman. "Bayesian model selection and model averaging." In: *Journal of mathematical psychology* 44.1 (2000), pp. 92–107.

[59] Yahoo Finance. *S&P 500 (^GSPC) — Quote and Historical Data*. Data source. 2025. URL: https://finance.yahoo.com/quote/%5EGSPC/ (visited on 11/09/2025).

[60] Walter Zucchini and Iain L MacDonald. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC, 2009.

Appendix

A.1 Code

Code used for this paper (and some for the keen reader) is available on this hyper link to a GitHub repository dedicated to this paper.

A.2 Derivations & Proofs

HMM and SSM We start by defining two key concepts to assist in most of the proofs: a directed acyclical graph and a parent of a r.v.[9].

Definition A.2.1. *Let $G = (V, E)$ be a directed graph, where V is a finite set of vertices (or nodes) and $E \subseteq V \times V$ is a set of directed edges, where an edge $(u, v) \in E$ indicates a directed link from u to v. The graph G is called a* directed acyclic graph (DAG) *if and only if it contains no directed cycles; that is, there do not exist distinct vertices $v_1, v_2, \ldots, v_k \in V$ such that $(v_i, v_{i+1}) \in E$ for all $i = 1, \ldots, k-1$, and $(v_k, v_1) \in E$. Equivalently, G is acyclic if there exists a topological ordering of the vertices $v_1, v_2, \ldots, v_n$ such that $(u, v) \in E \implies$ the index of u is less than that of v.*

Definition A.2.2. *Let $G = (V, E)$ be a directed acyclic graph (DAG), where $V = \{V_1, \ldots, V_n\}$ denotes the set of vertices (r.v.'s) and $E \subseteq V \times V$ denotes the set of directed edges. For a node $V_i \in V$, the* parent set *of V_i is defined as*

$$\mathrm{pa}(V_i) := \{\, V_j \in V : (V_j, V_i) \in E \,\}.$$

That is, V_j is said to be a parent *of V_i if and only if there exists a directed edge from V_j into V_i in the graph.*

The driving tool for any of the proofs is the following factorization for the joint distribution of the set of r.v.'s V_i $i \in \{1, \ldots, N\}$ in a directed acyclic graph

$$f_{\mathbf{V}^{(N)}}\left(\mathbf{v}^{(N)}\right) = \prod_{i=1}^{N} f_{V_i|\mathrm{pa}(V_i)}(v_i \mid \mathrm{pa}(v_i)), \tag{37}$$

where $\mathrm{pa}(V_i)$ denotes all the parents of V_i in the set $\{V_1, V_2, \ldots, V_N\}$. For example, consider our usual hidden Markov model setup such as that in Figure 7. The only parent of X_k is C_k and for $k = 2, 3, \ldots$ the only parent of C_k is C_{k-1} (obviously, C_1 has no parent). As an example, the joint distribution of $\mathbf{X}^{(t)}$ and $\mathbf{C}^{(t)}$ is therefore given by

$$f_{\mathbf{X}^{(t)},C_t}(\mathbf{x}^{(t)}, c) = \mathbb{P}(C_1) \prod_{k=2}^{t} \mathbb{P}(C_k \mid C_{k-1}) \prod_{k=1}^{t} f_{X_k,C_k}(x_k, c_k). \tag{38}$$

Lemma A.2.1. *For $t \in \mathbb{Z}^+$ and histories $\mathbf{X}^{(t)}$ and $\mathbf{C}^{(t)}$ we have that*

$$f_{\mathbf{X}^{(t+1)},C_t,C_{t+1}}(\mathbf{x}^{(t+1)}, c_t, c_{t+1}) = f_{\mathbf{X}^{(t)},C_t}(\mathbf{x}^{(t+1)}, c_t)\mathbb{P}\left(C_{t+1} \mid C_t\right) f_{X_{t+1},C_{t+1}}(x_{t+1}, c_{t+1})$$

[9]Proceeding in the appendix, we use the notation, that a draw of a r.v. (say C) is simply denoted by the lower caps of said r.v. (say c) and not necessarily the usual state values $i, j \in \mathcal{C}$.

Proof. By Equation 38 and the analogous expression for the expression $f_{\mathbf{X}^{(t+1)},\mathbf{C}^{(t+1)}}\left(\mathbf{x}^{(t+1)},\mathbf{c}^{(t+1)}\right)$ imply that

$$f_{\mathbf{X}^{(t+1)},\mathbf{C}^{(t+1)}}\left(\mathbf{x}^{(t+1)},\mathbf{c}^{(t+1)}\right)=\mathbb{P}(C_{t+1}\mid C_t)f_{\mathbf{X}^{(t)},\mathbf{C}^{(t)}}\left(\mathbf{x}^{(t)},\mathbf{c}^{(t)}\right)f_{X_{t+1}|C_{t+1}}(x_{t+1},c_{t+1})$$

Summing over $\mathbf{C}^{(t-1)}$ yields the desired result. □

Lemma A.2.2. *For* $t=1,2,\ldots,T-1$,

$$f_{\mathbf{X}_{t+1}^T|C_{t+1}}(\mathbf{x}_{t+1}^T\mid c_{t+1})=f_{X_{t+1}|C_{t+1}}(x_{t+1}\mid c_{t+1})f_{\mathbf{X}_{t+2}}(\mathbf{x}_{t+2})$$

Proof. The result follows by

$$\begin{aligned}f_{\mathbf{X}_{t+1}^T,\mathbf{C}_{t+1}^T}(\mathbf{X}_{t+1}^T,\mathbf{C}_{t+1}^T)&=f_{X_{t+1}|C_{t+1}}(x_{t+1}\mid c_{t+1})\left(\mathbb{P}\left(C_{t+1}\right)\prod_{k=t+2}^{T}\mathbb{P}\left(C_k\mid C_{k-1}\right)\prod_{k=t+2}^{T}f_{X_k|C_k}(x_k\mid c_k)\right)\\&=f_{X_{t+1}|C_{t+1}}(x_{t+1}\mid c_{t+1})f_{\mathbf{X}_{t+2}^T,\mathbf{C}_{t+1}^T}(\mathbf{x}_{t+2}^T,\mathbf{c}_{t+1}^T)\end{aligned}$$

and then summing over $\mathbf{C}_{t+2}^T$ and dividing by $\mathbb{P}\left(C_{t+1}\right)$. □

Lemma A.2.3. *For* $t=1,2,\ldots,T-1$,

$$f_{\mathbf{X}_{t+1}^T|C_{t+1}}(\mathbf{x}_{t+1}^T\mid c_{t+1})=f_{\mathbf{X}_{t+1}^T|C_{t+1},C_t}(\mathbf{x}_{t+1}^T\mid c_t,c_{t+1}).\quad(\dagger)$$

Proof. Simply rewrite the RHS of (†) to

$$\frac{1}{\mathbb{P}(C_t,C_{t+1})}\sum_{\mathbf{C}_{t+2}^T}f_{\mathbf{X}_{t+1}^T,\mathbf{C}_t^T}(\mathbf{x}_{t+1}^T,\mathbf{c}_t^T),$$

which by Equation 37 reduces to

$$\sum_{\mathbf{C}_{t+2}^T}\prod_{k=t+2}^{T}\mathbb{P}(C_k\mid C_{k-1})\prod_{k=t+1}^{T}f_{X_k|C_k}(x_k\mid c_k).$$

The LHS of (†) is

$$\frac{1}{\mathbb{P}(C_t)}\sum_{\mathbf{C}_{t+2}^T}f_{\mathbf{X}_{t+1}^T,\mathbf{C}_t^T}(\mathbf{x}_{t+1}^T,\mathbf{c}_t^T)=\sum_{\mathbf{C}_{t+2}^T}\prod_{k=t+2}^{T}\mathbb{P}(C_k\mid C_{k-1})\prod_{k=t+1}^{T}f_{X_k|C_k}(x_k\mid c_k),$$

which show that both sides reduce to the same expression. □

Lemma A.2.4. *For $r = 1, \dots, T$ and $i_r \in \{1, \dots, m\}$, the vectors defined by*

$$\alpha_1(i_1) \equiv h(i_1), \qquad \alpha_{r+1}(i_{r+1}) \equiv \sum_{i_r=1}^{m} \alpha_r(i_r)\, f_{r+1}(i_r, i_{r+1})$$

satisfy

$$\alpha_r(i_r) = \sum_{i_1=1}^{m} \cdots \sum_{i_{r-1}=1}^{m} h(i_1) \prod_{t=2}^{r} f_t(i_{t-1}, i_t). \quad (\dagger)$$

Proof. We proceed by induction on r.

Base case ($r = 1$). The product over an empty index set equals 1, so

$$\alpha_1(i_1) = h(i_1) = \sum_{\text{empty}} h(i_1) \cdot 1,$$

which is $(\dagger)$ for $r = 1$.

Inductive step. Assume $(\dagger)$ holds for some $r \in \{1, \dots, T-1\}$. Then

$$\begin{aligned}
\alpha_{r+1}(i_{r+1}) &= \sum_{i_r=1}^{m} \alpha_r(i_r)\, f_{r+1}(i_r, i_{r+1}) \\
&= \sum_{i_r=1}^{m} \left[\sum_{i_1=1}^{m} \cdots \sum_{i_{r-1}=1}^{m} h(i_1) \prod_{t=2}^{r} f_t(i_{t-1}, i_t) \right] f_{r+1}(i_r, i_{r+1}) \\
&= \sum_{i_1=1}^{m} \cdots \sum_{i_r=1}^{m} h(i_1) \prod_{t=2}^{r+1} f_t(i_{t-1}, i_t),
\end{aligned}$$

which is $(\dagger)$ with r replaced by $r+1$. By induction, the claim holds for all r and in particular for $r = T$:

$$\alpha_T(i_T) = \sum_{i_1=1}^{m} \cdots \sum_{i_{T-1}=1}^{m} h(i_1) \prod_{t=2}^{T} f_t(i_{t-1}, i_t).$$

□

Lemma A.2.5. *Let $\mathbf{F}_t$ be the $m \times m$ matrix with (i, j) entry $f_t(i, j)$ and let $\mathbf{1}_N$ denote the $m \times 1$ vector of ones. Then*

$$\mathbb{S} = \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} h(i_1) \prod_{t=2}^{T} f_t(i_{t-1}, i_t) = \sum_{i_T=1}^{m} \alpha_T(i_T) = \boldsymbol{\alpha}_T \mathbf{1}_N = \boldsymbol{\alpha}_1 \mathbf{F}_2 \mathbf{F}_3 \cdots \mathbf{F}_T \mathbf{1}_N.$$

Proof. The first equality is the definition of $\mathbb{S}$. The second follows from Lemma A.2.4 with $r = T$, which shows $\alpha_T(i_T)$ is the sum over all indices except i_T. Summing over i_T gives $\mathbb{S} = \sum_{i_T} \alpha_T(i_T) = \boldsymbol{\alpha}_T \mathbf{1}_N$. Finally, by construction $\boldsymbol{\alpha}_{r+1} = \boldsymbol{\alpha}_r \mathbf{F}_{r+1}$, hence $\boldsymbol{\alpha}_T = \boldsymbol{\alpha}_1 \mathbf{F}_2 \cdots \mathbf{F}_T$, yielding the formula. □

AR(1) process We prove a Lemma to help us rewrite the AR(1) process in a more convenient form under certain conditions.

Lemma A.2.6. *With $\rho \in \mathbb{R}$ and $\rho \neq 1$, then*

$$1 + \rho + \rho^2 + \ldots + \rho^n = \sum_{i=0}^{n} \rho^i = \left(1 - \rho^{n+1}\right)/(1 - \rho).$$

If moreover $|\rho| < 1$, $\rho^n \to 0$ as $n \to \infty$ and

$$\sum_{i=0}^{\infty} \rho^i = 1/(1 - \rho)$$

Proof. Let $S_n := \sum_{i=0}^{n} \rho^i$ with $\rho \in \mathbb{R}$ and $\rho \neq 1$. Then

$$(1 - \rho)S_n = \sum_{i=0}^{n} \rho^i - \sum_{i=0}^{n} \rho^{i+1} = \left(1 + \rho + \rho^2 + \ldots + \rho^n\right) - \left(\rho + \rho^2 + \ldots + \rho^{n+1}\right) = 1 - \rho^{n+1}.$$

Hence

$$S_n = \frac{1 - \rho^{n+1}}{1 - \rho},$$

which proves the finite–sum identity. If moreover $|\rho| < 1$, then $|\rho|^n \to 0$ as $n \to \infty$. Taking limits in the identity above yields

$$\sum_{i=0}^{\infty} \rho^i = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{1 - \rho^{n+1}}{1 - \rho} = \frac{1 - 0}{1 - \rho} = \frac{1}{1 - \rho}.$$

□

A.3 Figures

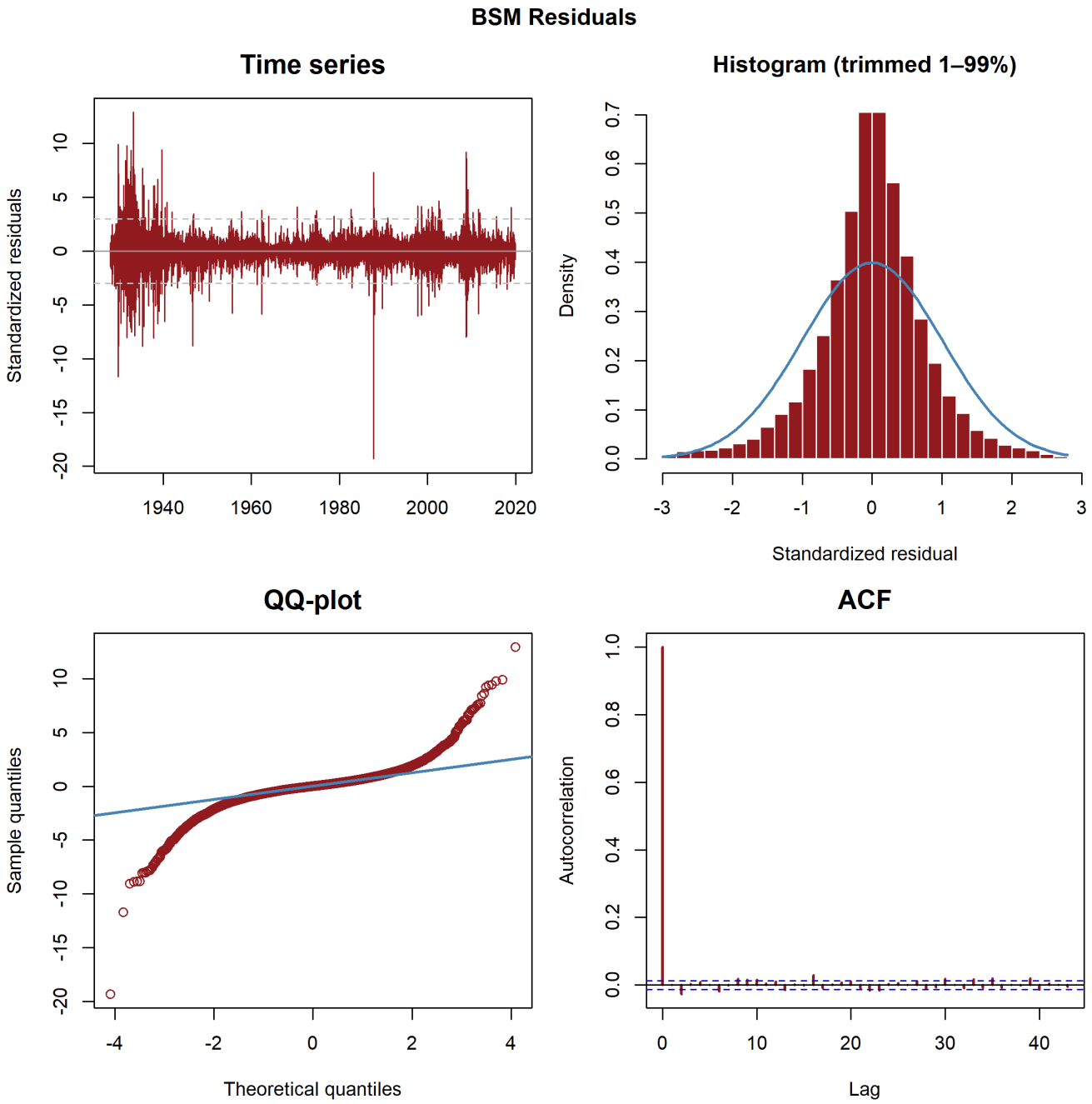


Figure A.3.1: *Standarized residuals for the BSM. The histogram is trimmed for the purpose of inspection.*

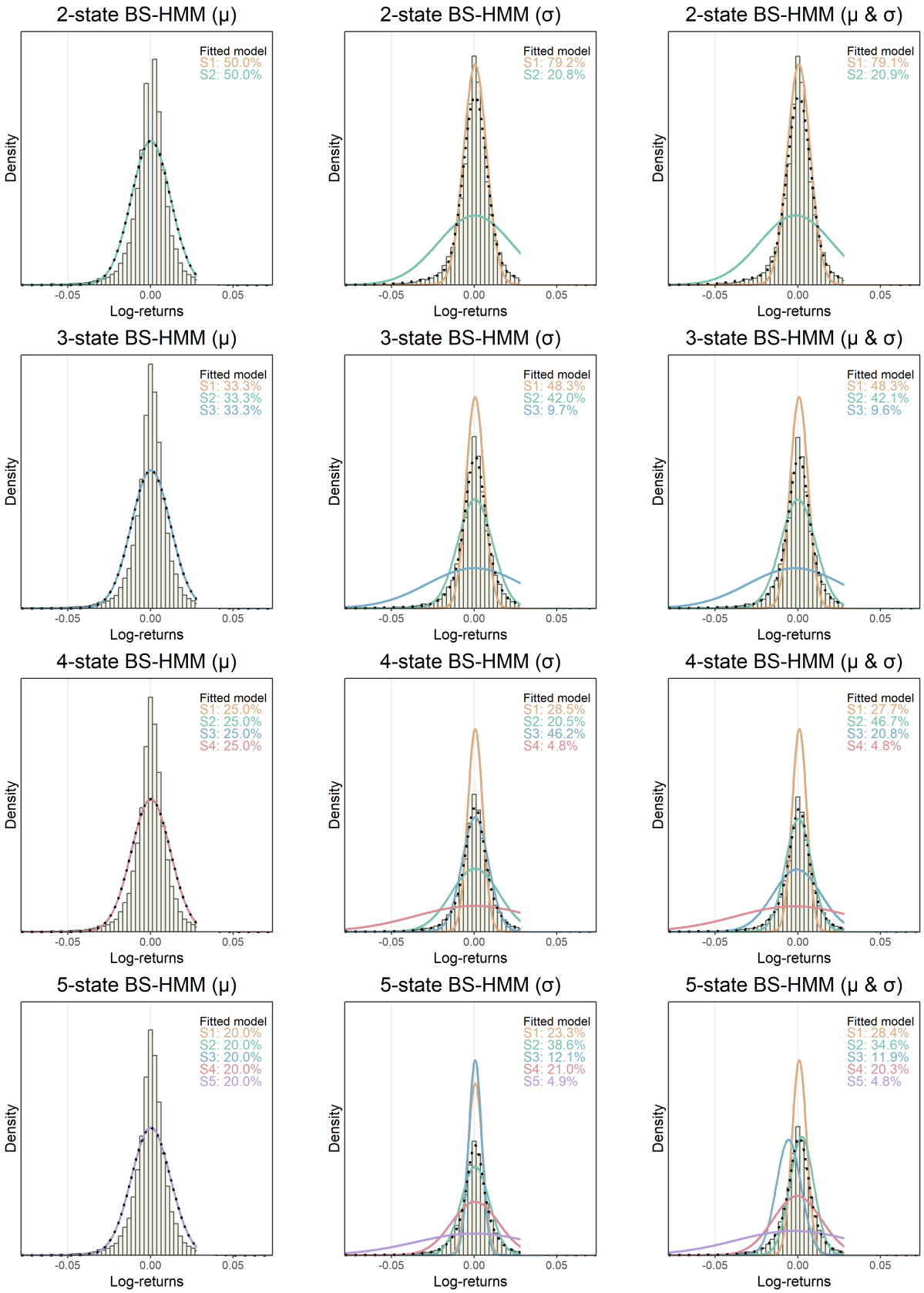


Figure A.3.2: *State-dependent density plots for the BS-HMMs.*

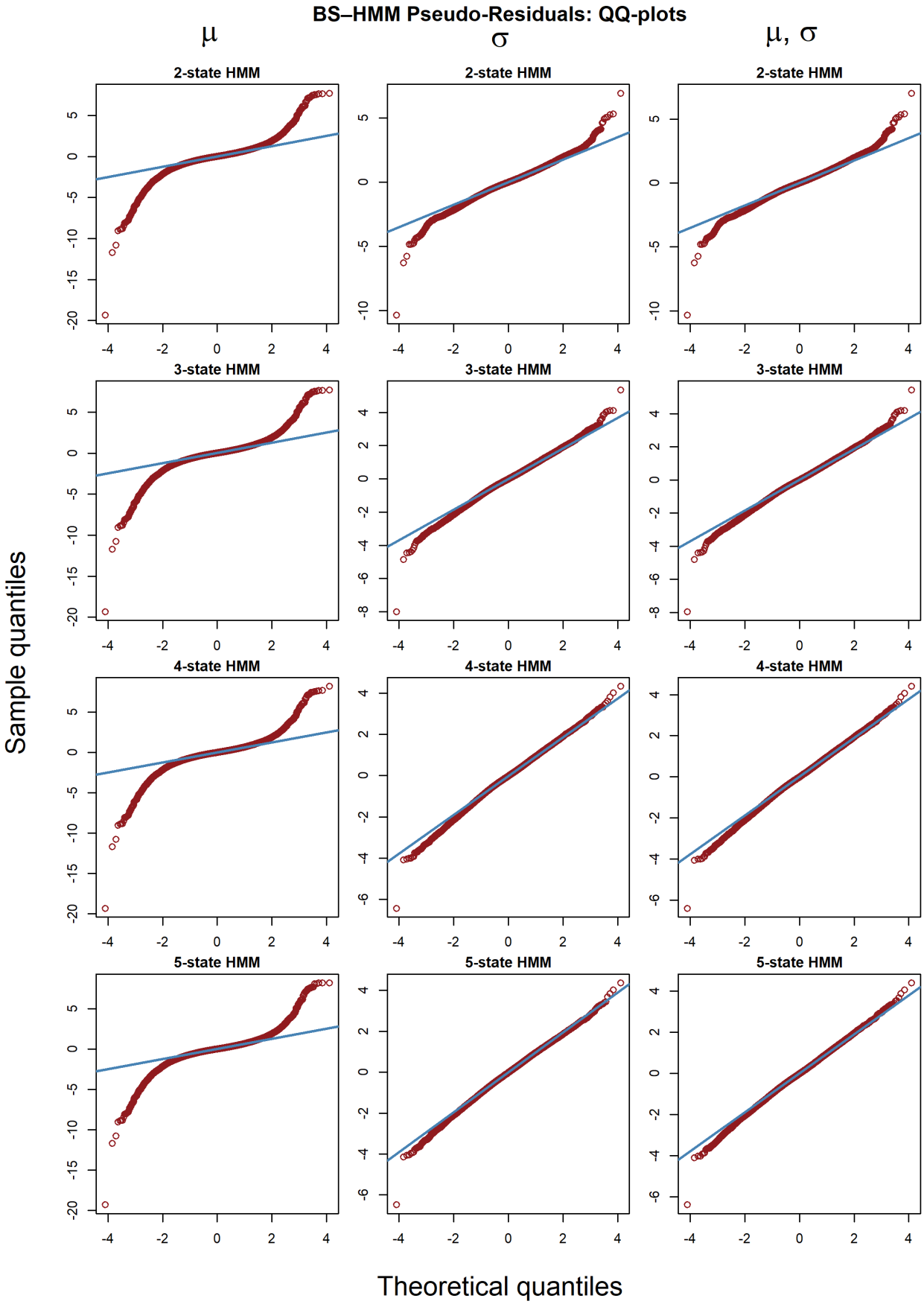


Figure A.3.3: *Q-Q-plot for the BS-HMMs.*

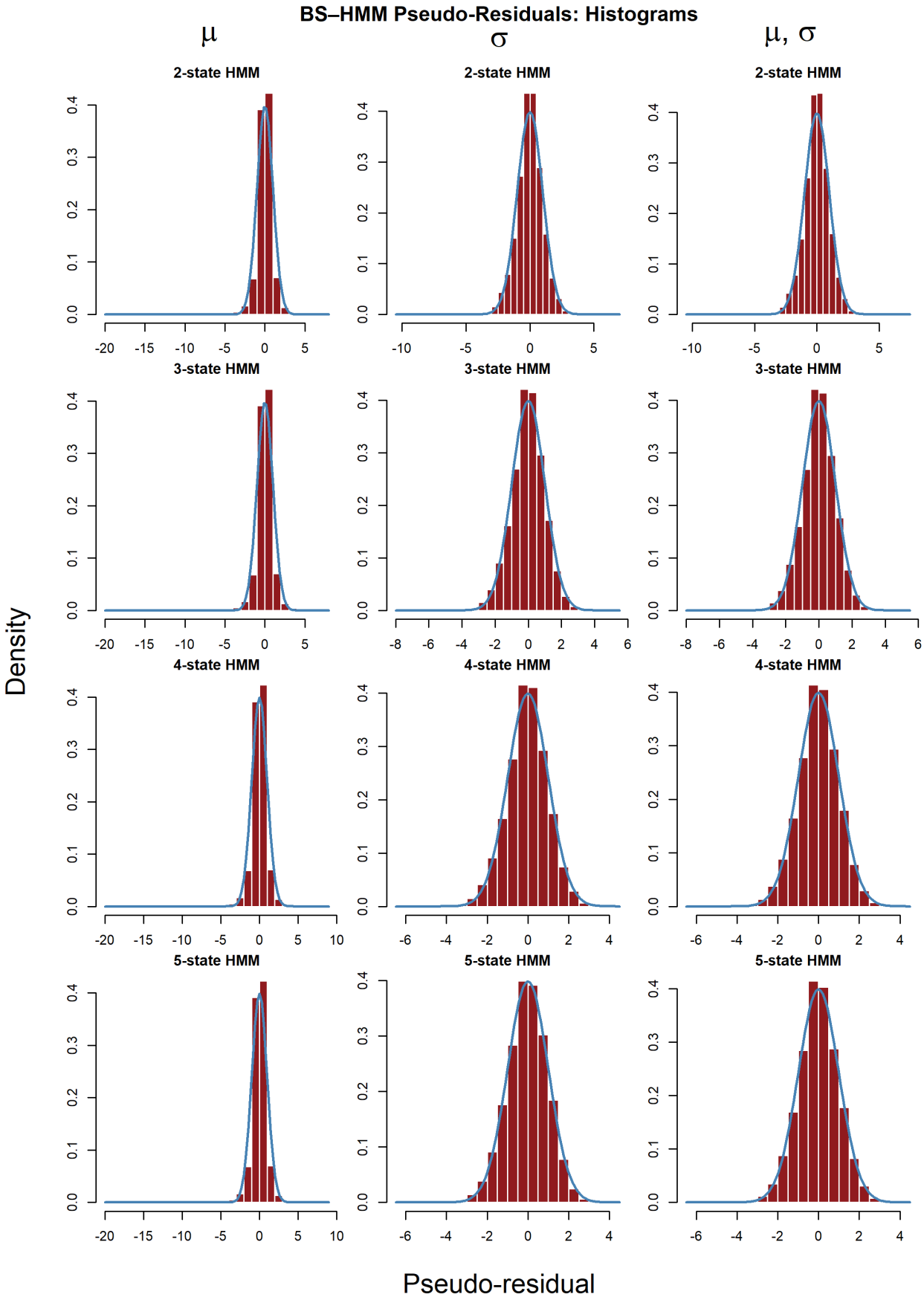


Figure A.3.4: *Histograms for the BS-HMMs. We trimmed the residuals for inspection*

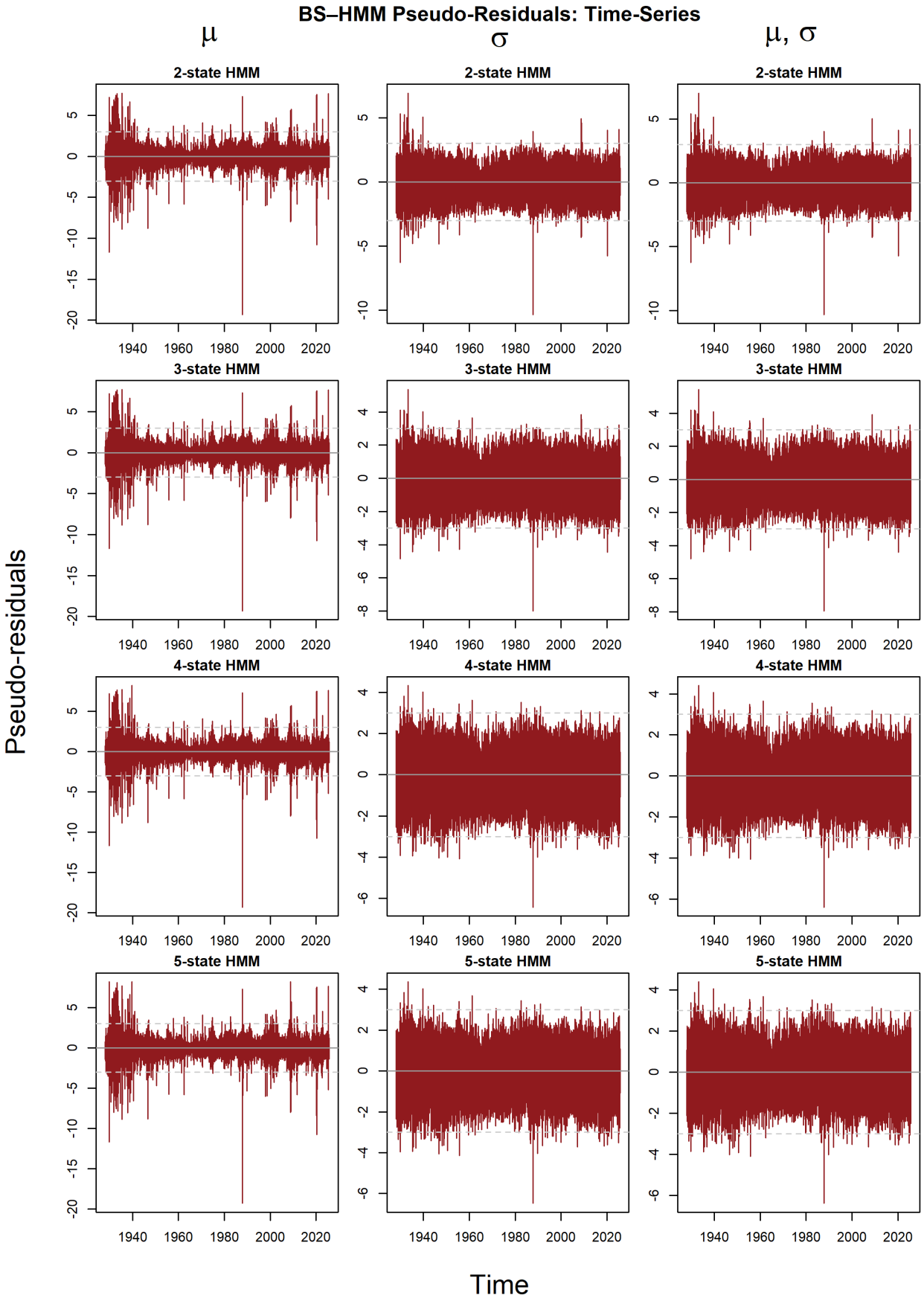


Figure A.3.5: *Histograms for the BS-HMMs. We trimmed the residuals for inspection*

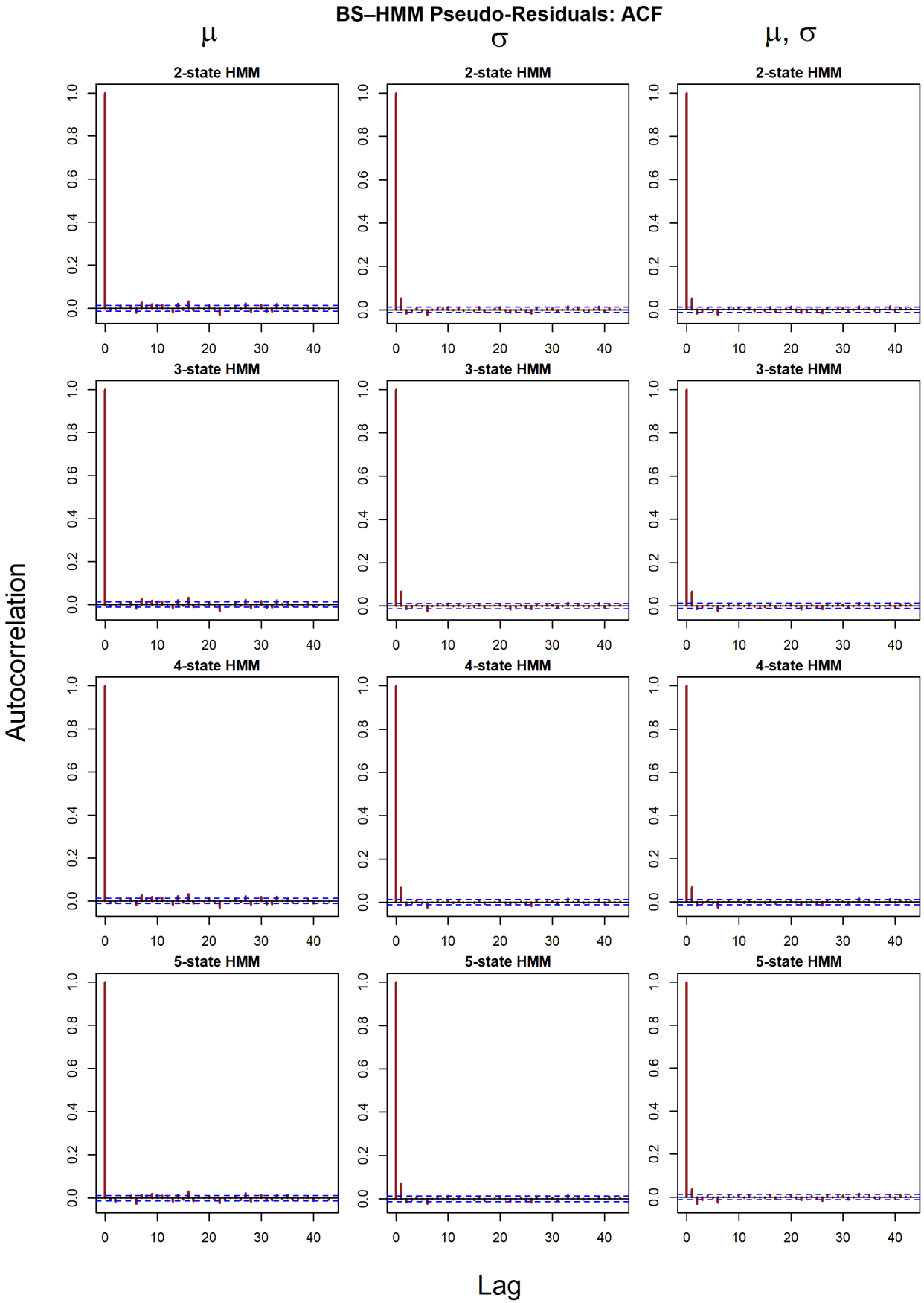


Figure A.3.6: *ACF plot for the BS-HMMs.*

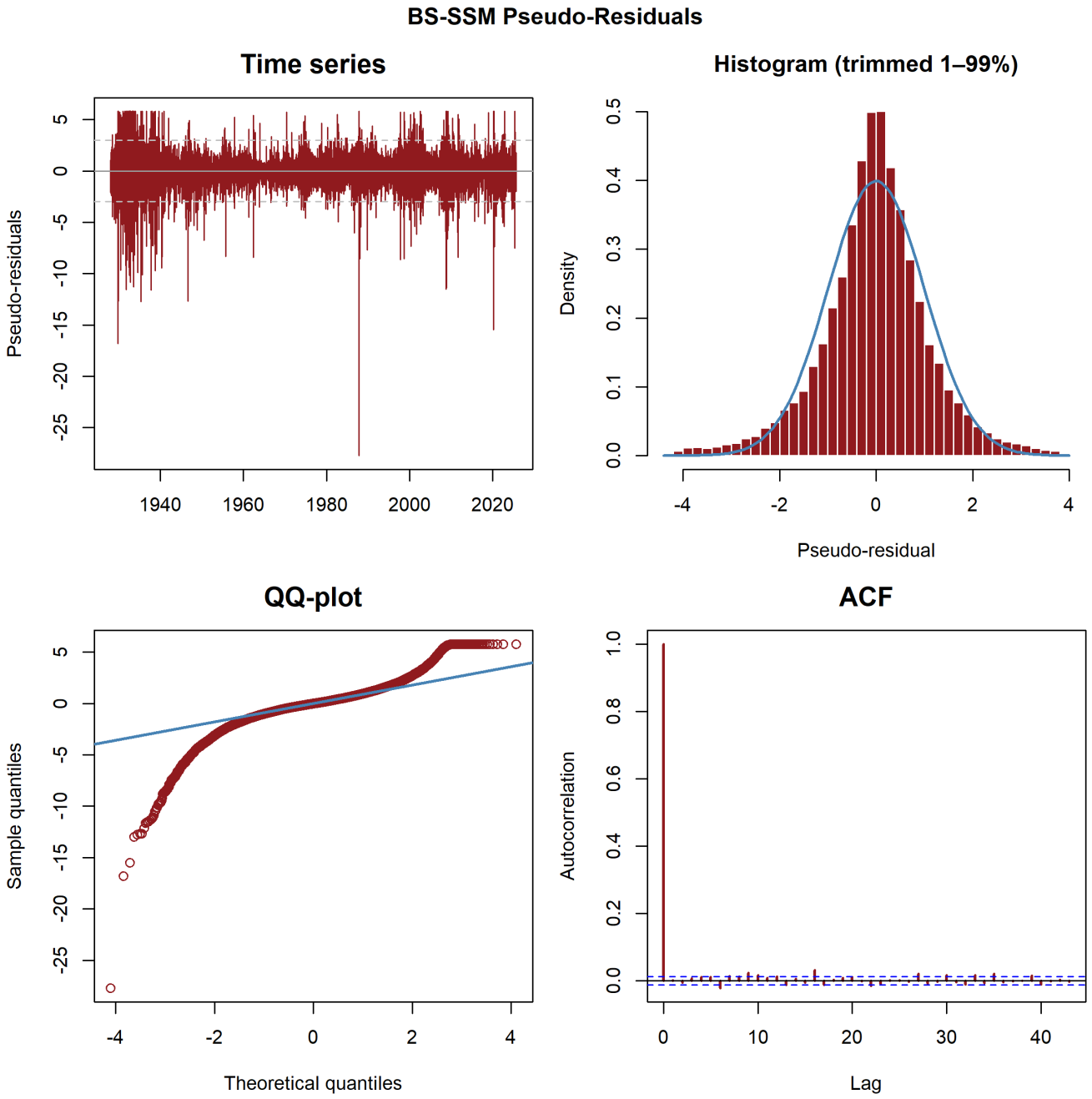


Figure A.3.7: *Pseudo-residuals for the BS-SSM. The histogram is trimmed for the purpose of inspection.*

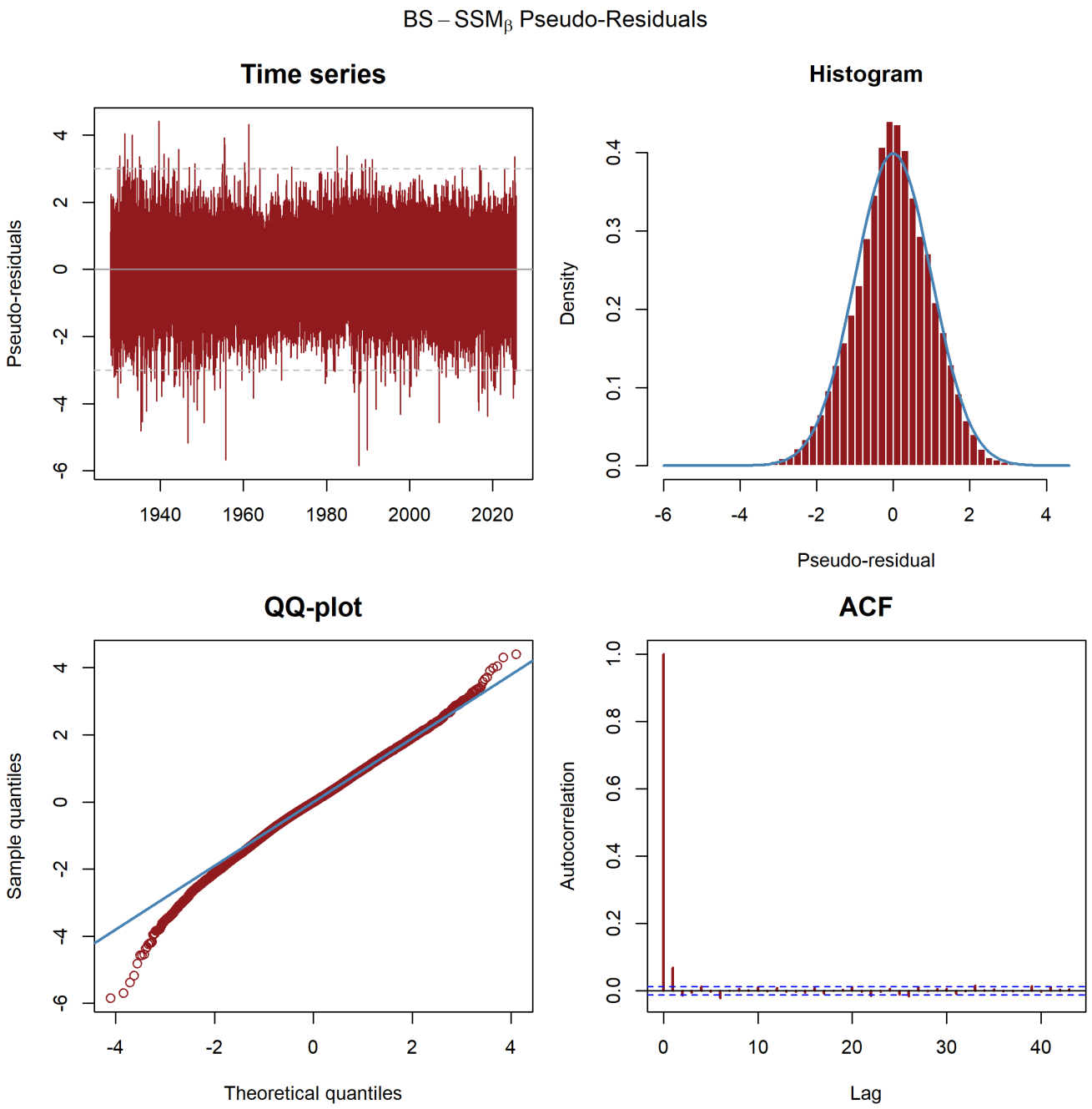


Figure A.3.8: *Pseudo-residuals for the BS-SSM$_\beta$.*

A.4 Tables

Black-Scholes Hidden Markov Model We list tables for the BS-HMMs below.

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.0746 (-0.1947, 0.3439)	0.13741
$\widehat{\mu}_{\text{cap},2}$	0.0745 (-0.1943, 0.3433)	0.13715
$\widehat{q}_1$	NA (—, —)	—
$\widehat{q}_2$	0.0340 (0.0322, 0.0359)	0.00092
$\widehat{\mu}_{\text{tot},1}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1085 (—, —)	—
$\widehat{\sigma}$	0.1883 (0.1866, 0.1900)	0.00088
$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.8807\ (0.6190) & 0.1193\ (0.6190) \\ 0.1191\ (0.5486) & 0.8809\ (0.5486) \end{pmatrix}$		
$\widehat{\boldsymbol{\delta}} = (0.4997\ (1.3079), 0.5003\ (1.3079))$		

Table A.4.1: *2-state BS-HMM with state-dependent drift $\mu_{cap,i}$ with common volatility. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.0723 (-0.3200, 0.4646)	0.20014
$\widehat{\mu}_{\text{cap},2}$	0.0783 (-0.3000, 0.4566)	0.19300
$\widehat{\mu}_{\text{cap},3}$	0.0730 (-0.3157, 0.4617)	0.19830
$\widehat{q}_1$	NA (—, —)	—
$\widehat{q}_2$	NA (—, —)	—
$\widehat{q}_3$	0.0340 (0.0322, 0.0359)	0.00092
$\widehat{\mu}_{\text{tot},1}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	0.1071 (—, —)	—
$\widehat{\sigma}$	0.1883 (0.1866, 0.1900)	0.00088

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.7869\ (0.1183) & 0.1065\ (0.3642) & 0.1066\ (—) \\ 0.1065\ (—) & 0.7870\ (—) & 0.1065\ (0.2975) \\ 0.1065\ (—) & 0.1065\ (—) & 0.7870\ (0.1004) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.3332\ (—), 0.3334\ (—), 0.3334\ (—))$$

Table A.4.2: *3-state BS-HMM with state-dependent drift $\mu_{cap,i}$ with common volatility. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.0716 (-0.4523, 0.5955)	0.26728
$\widehat{\mu}_{\text{cap},2}$	0.0770 (-0.4153, 0.5694)	0.25120
$\widehat{\mu}_{\text{cap},3}$	0.0770 (-0.4231, 0.5771)	0.25516
$\widehat{\mu}_{\text{cap},4}$	0.0725 (-0.4402, 0.5853)	0.26162
$\widehat{q}_1$	NA (—, —)	—
$\widehat{q}_2$	NA (—, —)	—
$\widehat{q}_3$	NA (—, —)	—
$\widehat{q}_4$	0.0340 (0.0322, 0.0359)	0.00092
$\widehat{\mu}_{\text{tot},1}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	0.1066 (—, —)	—
$\widehat{\sigma}$	0.1883 (0.1866, 0.1900)	0.00088

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.7112\ (—) & 0.0963\ (0.1594) & 0.0963\ (0.3092) & 0.0963\ (0.1551) \\ 0.0962\ (—) & 0.7112\ (—) & 0.0963\ (—) & 0.0963\ (—) \\ 0.0962\ (0.3498) & 0.0962\ (0.2067) & 0.7113\ (—) & 0.0963\ (0.3889) \\ 0.0962\ (0.3526) & 0.0963\ (0.1787) & 0.0963\ (0.1636) & 0.7113\ (—) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.2499\ (—), 0.2500\ (0.1264), 0.2500\ (—), 0.2501\ (0.1970))$$

Table A.4.3: *4-state BS-HMM with state-dependent drift $\mu_{cap,i}$ with common volatility. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.0714 (-0.5945, 0.7373)	0.33974
$\widehat{\mu}_{\text{cap},2}$	0.0758 (-0.5424, 0.6941)	0.31541
$\widehat{\mu}_{\text{cap},3}$	0.0773 (-0.5290, 0.6836)	0.30934
$\widehat{\mu}_{\text{cap},4}$	0.0760 (-0.5487, 0.7007)	0.31873
$\widehat{\mu}_{\text{cap},5}$	0.0722 (-0.5680, 0.7125)	0.32667
$\widehat{q}_1$	NA (—, —)	—
$\widehat{q}_2$	NA (—, —)	—
$\widehat{q}_3$	NA (—, —)	—
$\widehat{q}_4$	0.0340 (0.0322, 0.0359)	0.00092
$\widehat{q}_5$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},1}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	NA (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	0.1100 (—, —)	—
$\widehat{\mu}_{\text{tot},5}$	NA (—, —)	—
$\widehat{\sigma}$	0.1883 (0.1866, 0.1900)	0.00088

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.6487\ (—) & 0.0878\ (—) & 0.0878\ (—) & 0.0878\ (—) & 0.0878\ (—) \\ 0.0878\ (—) & 0.6488\ (—) & 0.0878\ (—) & 0.0878\ (—) & 0.0878\ (—) \\ 0.0878\ (0.1984) & 0.0878\ (—) & 0.6488\ (—) & 0.0878\ (—) & 0.0878\ (—) \\ 0.0878\ (0.0336) & 0.0878\ (—) & 0.0878\ (0.2075) & 0.6488\ (0.2634) & 0.0878\ (0.1643) \\ 0.0878\ (0.2050) & 0.0878\ (—) & 0.0878\ (—) & 0.0878\ (—) & 0.6488\ (—) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.2000\ (0.1697), 0.2000\ (—), 0.2000\ (0.2024), 0.2000\ (—), 0.2000\ (—))$$

Table A.4.4: *5-state BS-HMM with state-dependent drift $\mu_{cap,i}$ with common volatility. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap}}$	0.1443 (0.1182, 0.1703)	0.01329
$\widehat{q}_1$	0.0332 (0.0316, 0.0348)	0.00083
$\widehat{q}_2$	0.0374 (0.0335, 0.0413)	0.00201
$\widehat{\mu}_{\text{tot},1}$	0.1774 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1817 (—, —)	—
$\widehat{\sigma}_1$	0.1105 (0.1089, 0.1122)	0.00083
$\widehat{\sigma}_2$	0.3517 (0.3425, 0.3609)	0.00470
$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9888\ (0.0011) & 0.0112\ (0.0011) \\ 0.0427\ (0.0041) & 0.9573\ (0.0041) \end{pmatrix}$		
$\widehat{\boldsymbol{\delta}} = (0.7915\ (0.0167), 0.2085\ (0.0167))$		

Table A.4.5: *2-state BS-HMM with state-dependent volatility σ_i with common drift. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap}}$	0.1456 (0.1215, 0.1698)	0.01231
$\widehat{q}_1$	0.0323 (0.0305, 0.0341)	0.00091
$\widehat{q}_2$	0.0339 (0.0317, 0.0361)	0.00110
$\widehat{q}_3$	0.0439 (0.0382, 0.0496)	0.00290
$\widehat{\mu}_{\text{tot},1}$	0.1779 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1795 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	0.1896 (—, —)	—
$\widehat{\sigma}_1$	0.0865 (0.0843, 0.0886)	0.00108
$\widehat{\sigma}_2$	0.1675 (0.1629, 0.1721)	0.00237
$\widehat{\sigma}_3$	0.4554 (0.4388, 0.4721)	0.00850
$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9821\ (0.0019) & 0.0174\ (0.0020) & 0.0005\ (0.0005) \\ 0.0206\ (0.0022) & 0.9708\ (0.0026) & 0.0086\ (0.0013) \\ 0.0000\ (0.0000) & 0.0399\ (0.0054) & 0.9601\ (0.0054) \end{pmatrix}$		
$\widehat{\boldsymbol{\delta}} = (0.4833\ (0.0281), 0.4201\ (0.0236), 0.0966\ (0.0140))$		

Table A.4.6: *3-state BS-HMM with state-dependent volatility σ_i with common drift. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap}}$	0.1618 (0.1382, 0.1854)	0.01204
$\widehat{q}_1$	0.0312 (0.0292, 0.0333)	0.00104
$\widehat{q}_2$	0.0322 (0.0293, 0.0350)	0.00146
$\widehat{q}_3$	0.0347 (0.0328, 0.0366)	0.00098
$\widehat{q}_4$	0.0514 (0.0427, 0.0600)	0.00443
$\widehat{\mu}_{\text{tot},1}$	0.1931 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1940 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	0.1965 (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	0.2132 (—, —)	—
$\widehat{\sigma}_1$	0.0721 (0.0692, 0.0750)	0.00148
$\widehat{\sigma}_2$	0.2317 (0.2219, 0.2415)	0.00500
$\widehat{\sigma}_3$	0.1278 (0.1235, 0.1322)	0.00221
$\widehat{\sigma}_4$	0.5677 (0.5351, 0.6002)	0.01660

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9688\ (0.0040) & 0.0000\ (0.0000) & 0.0305\ (0.0040) & 0.0007\ (0.0004) \\ 0.0000\ (0.0000) & 0.9630\ (0.0039) & 0.0260\ (0.0034) & 0.0110\ (0.0021) \\ 0.0193\ (0.0028) & 0.0110\ (0.0015) & 0.9696\ (0.0031) & 0.0000\ (—) \\ 0.0000\ (0.0000) & 0.0521\ (0.0094) & 0.0000\ (0.0000) & 0.9479\ (0.0094) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.2850\ (0.0270), 0.2052\ (0.0221), 0.4622\ (0.0238), 0.0477\ (0.0098))$$

Table A.4.7: *4-state BS-HMM with state-dependent volatility σ_i with common drift. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap}}$	0.1578 (0.1346, 0.1810)	0.01184
$\widehat{q}_1$	0.0324 (0.0306, 0.0343)	0.00094
$\widehat{q}_2$	0.0320 (0.0285, 0.0354)	0.00177
$\widehat{q}_3$	0.0318 (0.0263, 0.0372)	0.00277
$\widehat{q}_4$	0.0343 (0.0321, 0.0365)	0.00112
$\widehat{q}_5$	0.0507 (0.0422, 0.0591)	0.00433
$\widehat{\mu}_{\text{tot},1}$	0.1902 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1898 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	0.1896 (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	0.1921 (—, —)	—
$\widehat{\mu}_{\text{tot},5}$	0.2085 (—, —)	—
$\widehat{\sigma}_1$	0.0707 (0.0678, 0.0736)	0.00147
$\widehat{\sigma}_2$	0.1375 (0.1314, 0.1435)	0.00309
$\widehat{\sigma}_3$	0.0623 (0.0509, 0.0737)	0.00582
$\widehat{\sigma}_4$	0.2274 (0.2175, 0.2374)	0.00506
$\widehat{\sigma}_5$	0.5623 (0.5301, 0.5945)	0.01642

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9757\ (0.0035) & 0.0000\ (0.0000) & 0.0238\ (0.0035) & 0.0000\ (0.0000) & 0.0005\ (0.0006) \\ 0.0147\ (0.0024) & 0.7156\ (0.0508) & 0.2580\ (0.0500) & 0.0111\ (0.0017) & 0.0006\ (0.0006) \\ 0.0000\ (—) & 0.9066\ (0.0748) & 0.0933\ (0.0748) & 0.0000\ (0.0000) & 0.0000\ (—) \\ 0.0000\ (0.0000) & 0.0000\ (—) & 0.0222\ (0.0030) & 0.9672\ (0.0036) & 0.0106\ (0.0021) \\ 0.0000\ (0.0000) & 0.0000\ (0.0000) & 0.0000\ (0.0000) & 0.0524\ (0.0092) & 0.9476\ (0.0092) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.2330\ (0.0279), 0.3864\ (0.0270), 0.1212\ (0.0217), 0.2102\ (0.0237), 0.0493\ (0.0101))$$

Table A.4.8: *5-state BS-HMM with state-dependent volatility σ_i with common drift. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.1587 (0.1319, 0.1854)	0.01363
$\widehat{\mu}_{\text{cap},2}$	-0.2431 (-0.4030, -0.0832)	0.08158
$\widehat{q}_1$	0.0332 (0.0316, 0.0348)	0.00083
$\widehat{q}_2$	0.0374 (0.0335, 0.0414)	0.00201
$\widehat{\mu}_{\text{tot},1}$	0.1918 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	-0.2057 (—, —)	—
$\widehat{\sigma}_1$	0.1104 (0.1088, 0.1120)	0.00083
$\widehat{\sigma}_2$	0.3502 (0.3410, 0.3593)	0.00466
$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9887\ (0.0011) & 0.0113\ (0.0011) \\ 0.0428\ (0.0041) & 0.9572\ (0.0041) \end{pmatrix}$		
$\widehat{\boldsymbol{\delta}} = (0.7908\ (0.0167), 0.2092\ (0.0167))$		

Table A.4.9: *2-state BS-HMM with state-dependent drift $\mu_{cap,i}$ and volatility σ_i. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.1824 (0.1533, 0.2116)	0.01486
$\widehat{\mu}_{\text{cap},2}$	0.0425 (-0.0155, 0.1004)	0.02956
$\widehat{\mu}_{\text{cap},3}$	-0.3245 (-0.6299, -0.0190)	0.15583
$\widehat{q}_1$	0.0323 (0.0305, 0.0341)	0.00091
$\widehat{q}_2$	0.0339 (0.0317, 0.0360)	0.00110
$\widehat{q}_3$	0.0439 (0.0382, 0.0496)	0.00289
$\widehat{\mu}_{\text{tot},1}$	0.2147 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.0764 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	-0.2806 (—, —)	—
$\widehat{\sigma}_1$	0.0862 (0.0840, 0.0883)	0.00110
$\widehat{\sigma}_2$	0.1675 (0.1628, 0.1722)	0.00239
$\widehat{\sigma}_3$	0.4543 (0.4376, 0.4710)	0.00850
$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9812\ (0.0021) & 0.0183\ (0.0021) & 0.0005\ (0.0004) \\ 0.0216\ (0.0024) & 0.9699\ (0.0027) & 0.0086\ (0.0013) \\ 0.0000\ (—) & 0.0398\ (0.0054) & 0.9602\ (0.0054) \end{pmatrix}$		
$\widehat{\boldsymbol{\delta}} = (0.4827\ (0.0278), 0.4209\ (0.0233), 0.0964\ (0.0140))$		

Table A.4.10: *3-state BS-HMM with state-dependent drift $\mu_{cap,i}$ and volatility σ_i. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.2346 (0.1990, 0.2703)	0.01820
$\widehat{\mu}_{\text{cap},2}$	0.1062 (0.0621, 0.1502)	0.02246
$\widehat{\mu}_{\text{cap},3}$	-0.1031 (-0.2193, 0.0132)	0.05930
$\widehat{\mu}_{\text{cap},4}$	-0.3818 (-0.9260, 0.1623)	0.27763
$\widehat{q}_1$	0.0310 (0.0289, 0.0330)	0.00104
$\widehat{q}_2$	0.0348 (0.0329, 0.0368)	0.00099
$\widehat{q}_3$	0.0321 (0.0293, 0.0349)	0.00145
$\widehat{q}_4$	0.0512 (0.0426, 0.0599)	0.00442
$\widehat{\mu}_{\text{tot},1}$	0.2656 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.1410 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	-0.0710 (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	-0.3306 (—, —)	—
$\widehat{\sigma}_1$	0.0707 (0.0680, 0.0734)	0.00139
$\widehat{\sigma}_2$	0.1270 (0.1229, 0.1311)	0.00210
$\widehat{\sigma}_3$	0.2301 (0.2205, 0.2396)	0.00487
$\widehat{\sigma}_4$	0.5651 (0.5329, 0.5974)	0.01646

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9649\ (0.0044) & 0.0341\ (0.0045) & 0.0002\ (0.0006) & 0.0007\ (0.0005) \\ 0.0207\ (0.0030) & 0.9683\ (0.0033) & 0.0110\ (0.0015) & 0.0000\ (—) \\ 0.0000\ (0.0000) & 0.0259\ (0.0033) & 0.9632\ (0.0039) & 0.0109\ (0.0021) \\ 0.0000\ (0.0000) & 0.0000\ (0.0000) & 0.0517\ (0.0094) & 0.9483\ (0.0094) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.2766\ (0.0254), 0.4675\ (0.0233), 0.2080\ (0.0220), 0.0479\ (0.0099))$$

Table A.4.11: *4-state BS-HMM with state-dependent drift $\mu_{cap,i}$ and volatility σ_i. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\mu}_{\text{cap},1}$	0.2181 (0.1833, 0.2529)	0.01777
$\widehat{\mu}_{\text{cap},2}$	0.6049 (0.4268, 0.7831)	0.09089
$\widehat{\mu}_{\text{cap},3}$	-1.3869 (-1.7642, -1.0097)	0.19247
$\widehat{\mu}_{\text{cap},4}$	-0.0672 (-0.1861, 0.0518)	0.06068
$\widehat{\mu}_{\text{cap},5}$	-0.3986 (-0.9461, 0.1488)	0.27930
$\widehat{q}_1$	0.0318 (0.0298, 0.0338)	0.00101
$\widehat{q}_2$	0.0353 (0.0329, 0.0376)	0.00119
$\widehat{q}_3$	0.0363 (0.0327, 0.0399)	0.00185
$\widehat{q}_4$	0.0331 (0.0307, 0.0354)	0.00121
$\widehat{q}_5$	0.0512 (0.0425, 0.0599)	0.00444
$\widehat{\mu}_{\text{tot},1}$	0.2499 (—, —)	—
$\widehat{\mu}_{\text{tot},2}$	0.6402 (—, —)	—
$\widehat{\mu}_{\text{tot},3}$	-1.3506 (—, —)	—
$\widehat{\mu}_{\text{tot},4}$	-0.0341 (—, —)	—
$\widehat{\mu}_{\text{tot},5}$	-0.3474 (—, —)	—
$\widehat{\sigma}_1$	0.0702 (0.0676, 0.0729)	0.00136
$\widehat{\sigma}_2$	0.1156 (0.1111, 0.1202)	0.00233
$\widehat{\sigma}_3$	0.1186 (0.1103, 0.1270)	0.00425
$\widehat{\sigma}_4$	0.2314 (0.2218, 0.2410)	0.00489
$\widehat{\sigma}_5$	0.5665 (0.5340, 0.5990)	0.01659

$$\widehat{\mathbf{\Gamma}} = \begin{pmatrix} 0.9626\ (0.0047) & 0.0000\ (0.0000) & 0.0368\ (0.0047) & 0.0000\ (0.0000) & 0.0006\ (0.0005) \\ 0.0307\ (0.0057) & 0.8341\ (0.0345) & 0.1349\ (0.0310) & 0.0000\ (—) & 0.0004\ (0.0006) \\ 0.0000\ (—) & 0.4388\ (0.0474) & 0.5190\ (0.0526) & 0.0422\ (0.0093) & 0.0000\ (—) \\ 0.0000\ (0.0000) & 0.0261\ (0.0034) & 0.0000\ (0.0000) & 0.9631\ (0.0040) & 0.0109\ (0.0021) \\ 0.0000\ (0.0000) & 0.0000\ (0.0000) & 0.0000\ (0.0000) & 0.0525\ (0.0094) & 0.9475\ (0.0094) \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = (0.2842\ (0.0244), 0.3461\ (0.0313), 0.1188\ (0.0245), 0.2033\ (0.0214), 0.0476\ (0.0097))$$

Table A.4.12: *5-state BS-HMM with state-dependent drift $\mu_{cap,i}$ and volatility σ_i. State-dependent parameter estimates (annualised) with 95% confidence intervals based on the inverse Hessian of the minimised log-likelihood on the working scale, transformed to the natural scale by the delta method. The capital-gains drift $\mu_{cap,i}$ and dividend yield q_i combine to give the total-return drift $\mu_{tot,i} = \mu_{cap,i} + q_i$ in each state. The bottom block reports the estimated transition matrix $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$; entries are reported as estimate (asymptotic standard error). Non-finite components are shown as dashes, while finite components are retained.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_\varepsilon$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	-6.0032 (-6.0032, -6.0032)	0.000000
$\widehat{\mu}_{\text{tot}}$	-5.9692 (—, —)	—
$\widehat{\sigma}$	0.2591 (0.2591, 0.2591)	0.000000

Table A.4.13: *BS-SSM using an $m = 20$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9990 (0.9984, 0.9996)	0.000304
$\widehat{\sigma}_\varepsilon$	0.0144 (0.0131, 0.0156)	0.000652
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1112 (0.0756, 0.1468)	0.018159
$\widehat{\mu}_{\text{tot}}$	0.1452 (—, —)	—
$\widehat{\sigma}$	0.1735 (0.1711, 0.1759)	0.001205

Table A.4.14: *BS-SSM using an $m = 70$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9990 (0.9984, 0.9996)	0.000307
$\widehat{\sigma}_\varepsilon$	0.0145 (0.0132, 0.0157)	0.000652
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1111 (0.0756, 0.1467)	0.018154
$\widehat{\mu}_{\text{tot}}$	0.1452 (—, —)	—
$\widehat{\sigma}$	0.1735 (0.1711, 0.1758)	0.001203

Table A.4.15: *BS-SSM using an $m = 100$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9990 (0.9984, 0.9996)	0.000309
$\widehat{\sigma}_\varepsilon$	0.0145 (0.0132, 0.0158)	0.000653
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1111 (0.0755, 0.1467)	0.018150
$\widehat{\mu}_{\text{tot}}$	0.1452 (—, —)	—
$\widehat{\sigma}$	0.1734 (0.1711, 0.1758)	0.001202

Table A.4.16: *BS-SSM using an $m = 200$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9943 (0.9927, 0.9959)	0.000820
$\widehat{\sigma}_\varepsilon$	0.0522 (0.0487, 0.0558)	0.001818
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1511 (0.1161, 0.1861)	0.017838
$\widehat{\mu}_{\text{tot}}$	0.1851 (—, —)	—
$\widehat{\sigma}$	0.1687 (0.1646, 0.1727)	0.002090

Table A.4.17: *BS-SSM using an $m = 40$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9943 (0.9927, 0.9959)	0.000823
$\widehat{\sigma}_\varepsilon$	0.0524 (0.0488, 0.0559)	0.001820
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1509 (0.1160, 0.1858)	0.017822
$\widehat{\mu}_{\text{tot}}$	0.1850 (—, —)	—
$\widehat{\sigma}$	0.1685 (0.1644, 0.1726)	0.002079

Table A.4.18: *BS-SSM using an $m = 70$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9943 (0.9927, 0.9959)	0.000824
$\widehat{\sigma}_\varepsilon$	0.0524 (0.0488, 0.0559)	0.001821
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1509 (0.1159, 0.1858)	0.017818
$\widehat{\mu}_{\text{tot}}$	0.1849 (—, —)	—
$\widehat{\sigma}$	0.1685 (0.1644, 0.1725)	0.002076

Table A.4.19: *BS-SSM using an $m = 100$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9943 (0.9927, 0.9959)	0.000825
$\widehat{\sigma}_\varepsilon$	0.0524 (0.0488, 0.0560)	0.001821
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1508 (0.1159, 0.1858)	0.017816
$\widehat{\mu}_{\text{tot}}$	0.1849 (—, —)	—
$\widehat{\sigma}$	0.1685 (0.1644, 0.1725)	0.002074

Table A.4.20: *BS-SSM using an $m = 200$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (—, —)	—
$\widehat{\sigma}_\varepsilon$	0.0000 (—, —)	—
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0409 (—, —)	—
$\widehat{\mu}_{\text{tot}}$	0.0749 (—, —)	—
$\widehat{\sigma}$	0.1418 (—, —)	—

Table A.4.21: *BS-SSM using an $m = 20$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9845 (0.9814, 0.9876)	0.001586
$\widehat{\sigma}_\varepsilon$	0.0921 (0.0852, 0.0990)	0.003499
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1425 (0.1126, 0.1724)	0.015252
$\widehat{\mu}_{\text{tot}}$	0.1766 (—, —)	—
$\widehat{\sigma}$	0.1351 (0.1257, 0.1445)	0.004802

Table A.4.22: *BS-SSM using an $m = 40$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9845 (0.9814, 0.9876)	0.001585
$\widehat{\sigma}_\varepsilon$	0.0921 (0.0852, 0.0989)	0.003493
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1427 (0.1128, 0.1726)	0.015266
$\widehat{\mu}_{\text{tot}}$	0.1767 (—, —)	—
$\widehat{\sigma}$	0.1353 (0.1259, 0.1447)	0.004797

Table A.4.23: *BS-SSM using an $m = 70$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9845 (0.9814, 0.9876)	0.001585
$\widehat{\sigma}_\varepsilon$	0.0921 (0.0852, 0.0989)	0.003492
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1427 (0.1128, 0.1726)	0.015269
$\widehat{\mu}_{\text{tot}}$	0.1768 (—, —)	—
$\widehat{\sigma}$	0.1353 (0.1259, 0.1447)	0.004796

Table A.4.24: *BS-SSM using an $m = 100$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9845 (0.9814, 0.9876)	0.001585
$\widehat{\sigma}_{\varepsilon}$	0.0921 (0.0852, 0.0989)	0.003493
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1427 (0.1128, 0.1727)	0.015271
$\widehat{\mu}_{\text{tot}}$	0.1768 (—, —)	—
$\widehat{\sigma}$	0.1353 (0.1259, 0.1447)	0.004796

Table A.4.25: *BS-SSM using an $m = 200$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0001 (0.0001, 0.0001)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0800 (0.0435, 0.1165)	0.018602
$\widehat{\mu}_{\text{tot}}$	0.1141 (—, —)	—
$\widehat{\sigma}$	0.1637 (0.1637, 0.1637)	0.000000

Table A.4.26: *BS-SSM using an $m = 20$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	7.1856 (7.1856, 7.1856)	0.000000
$\widehat{\mu}_{\text{tot}}$	7.2197 (—, —)	—
$\widehat{\sigma}$	0.4511 (0.4511, 0.4511)	0.000000

Table A.4.27: *BS-SSM using an $m = 40$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9842 (0.9810, 0.9873)	0.001614
$\widehat{\sigma}_{\varepsilon}$	0.0933 (0.0862, 0.1003)	0.003590
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1390 (0.1097, 0.1684)	0.014977
$\widehat{\mu}_{\text{tot}}$	0.1731 (—, —)	—
$\widehat{\sigma}$	0.1314 (0.1214, 0.1414)	0.005124

Table A.4.28: *BS-SSM using an $m = 70$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9842 (0.9810, 0.9873)	0.001614
$\widehat{\sigma}_{\varepsilon}$	0.0933 (0.0862, 0.1003)	0.003590
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1390 (0.1097, 0.1684)	0.014976
$\widehat{\mu}_{\text{tot}}$	0.1731 (—, —)	—
$\widehat{\sigma}$	0.1314 (0.1214, 0.1414)	0.005123

Table A.4.29: *BS-SSM using an $m = 100$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9842 (0.9810, 0.9873)	0.001614
$\widehat{\sigma}_{\varepsilon}$	0.0933 (0.0862, 0.1003)	0.003590
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1390 (0.1097, 0.1684)	0.014976
$\widehat{\mu}_{\text{tot}}$	0.1731 (—, —)	—
$\widehat{\sigma}$	0.1314 (0.1214, 0.1414)	0.005123

Table A.4.30: *BS-SSM using an $m = 200$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9983 (0.9983, 0.9983)	0.000003
$\widehat{\sigma}_{\varepsilon}$	0.0002 (0.0002, 0.0002)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	-2.7355 (-3.5777, -1.8933)	0.429717
$\widehat{\mu}_{\text{tot}}$	-2.7015 (—, —)	—
$\widehat{\sigma}$	3.0458 (2.7486, 3.3430)	0.151629

Table A.4.31: *BS-SSM using an $m = 40$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9994 (0.9994, 0.9994)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0023 (0.0023, 0.0024)	0.000018
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0550 (0.0492, 0.0609)	0.002972
$\widehat{\mu}_{\text{tot}}$	0.0891 (—, —)	—
$\widehat{\sigma}$	0.0276 (0.0276, 0.0276)	0.000000

Table A.4.32: *BS-SSM using an $m = 70$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9842 (0.9810, 0.9873)	0.001614
$\widehat{\sigma}_\varepsilon$	0.0933 (0.0862, 0.1003)	0.003590
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1391 (0.1097, 0.1684)	0.014977
$\widehat{\mu}_{\text{tot}}$	0.1731 (—, —)	—
$\widehat{\sigma}$	0.1314 (0.1214, 0.1414)	0.005123

Table A.4.33: *BS-SSM using an $m = 100$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9842 (0.9810, 0.9873)	0.001614
$\widehat{\sigma}_\varepsilon$	0.0933 (0.0862, 0.1003)	0.003590
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1391 (0.1097, 0.1684)	0.014976
$\widehat{\mu}_{\text{tot}}$	0.1731 (—, —)	—
$\widehat{\sigma}$	0.1314 (0.1214, 0.1414)	0.005123

Table A.4.34: *BS-SSM using an $m = 200$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Black-Scholes Continuous State Space Model

Black-Scholes Continuous State Space Beta Model Note that standard errors of the unidentifiable parameters can not be used as is.

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9999 (0.9999, 0.9999)	0.000003
$\widehat{\sigma}_\varepsilon$	0.0001 (0.0001, 0.0001)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0815 (0.0815, 0.0815)	0.000001
$\widehat{\mu}_{\text{tot}}$	0.1156 (—, —)	—
$\widehat{\sigma}$	0.0204 (0.0204, 0.0204)	0.000000
$\widehat{\beta}_\mu$	-1.5964 (-1.6129, -1.5799)	0.008424
$\widehat{\beta}_\sigma$	29.7184 (29.7184, 29.7184)	0.000000

Table A.4.35: *BS-SSM$_\beta$ using an $m = 20$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9999 (0.9999, 0.9999)	0.000000
$\widehat{\sigma}_\varepsilon$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	-0.2213 (-0.3395, -0.1032)	0.060291
$\widehat{\mu}_{\text{tot}}$	-0.1873 (—, —)	—
$\widehat{\sigma}$	0.0494 (0.0494, 0.0494)	0.000000
$\widehat{\beta}_\mu$	-2.2842 (-5.2731, 0.7047)	1.524932
$\widehat{\beta}_\sigma$	34.0251 (34.0251, 34.0251)	0.000000

Table A.4.36: *BS-SSM$_\beta$ using an $m = 40$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_\varepsilon$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0244 (0.0244, 0.0244)	0.000000
$\widehat{\mu}_{\text{tot}}$	0.0585 (—, —)	—
$\widehat{\sigma}$	0.2005 (0.2005, 0.2005)	0.000000
$\widehat{\beta}_\mu$	6.2974 (6.2974, 6.2974)	0.000000
$\widehat{\beta}_\sigma$	21.3210 (21.3210, 21.3210)	0.000000

Table A.4.37: *BS-SSM$_\beta$ using an $m = 70$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_\varepsilon$	0.0130 (0.0130, 0.0130)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017520
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004942
$\widehat{\beta}_\mu$	-1.4539 (-1.4539, -1.4539)	0.000000
$\widehat{\beta}_\sigma$	7.5824 (7.5824, 7.5824)	0.000000

Table A.4.38: *BS-SSM$_\beta$ using an $m = 100$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0130 (-0.3410, 0.3670)	0.180599
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0546, 0.1272)	0.018539
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004949
$\widehat{\beta}_{\mu}$	-1.4547 (-40.8457, 37.9362)	20.097417
$\widehat{\beta}_{\sigma}$	7.5868 (-199.0721, 214.2457)	105.438209

Table A.4.39: *BS-SSM$_{\beta}$ using an $m = 200$ point grid and truncation $b_{\max} = 0.5$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9990 (0.9990, 0.9990)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0001 (0.0001, 0.0001)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	-0.5994 (-0.8299, -0.3690)	0.117563
$\widehat{\mu}_{\text{tot}}$	-0.5654 (—, —)	—
$\widehat{\sigma}$	0.1062 (0.0249, 0.1875)	0.041498
$\widehat{\beta}_{\mu}$	0.0166 (-4.5620, 4.5952)	2.336028
$\widehat{\beta}_{\sigma}$	3.3083 (-12.0061, 18.6227)	7.813446

Table A.4.40: *BS-SSM$_{\beta}$ using an $m = 20$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	-5.2982 (-5.9325, -4.6640)	0.323609
$\widehat{\mu}_{\text{tot}}$	-5.2642 (—, —)	—
$\widehat{\sigma}$	28.8391 (28.0949, 29.5833)	0.379718
$\widehat{\beta}_{\mu}$	-37.7873 (-39.7743, -35.8002)	1.013803
$\widehat{\beta}_{\sigma}$	12.5466 (12.4915, 12.6017)	0.028107

Table A.4.41: *BS-SSM$_{\beta}$ using an $m = 40$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0365 (0.0365, 0.0365)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017531
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.5183 (-0.5183, -0.5183)	0.000000
$\widehat{\beta}_{\sigma}$	2.7028 (2.7028, 2.7028)	0.000000

Table A.4.42: *BS-SSM$_{\beta}$ using an $m = 70$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0365 (0.0365, 0.0365)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017528
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.5186 (-0.5186, -0.5186)	0.000000
$\widehat{\beta}_{\sigma}$	2.7043 (2.7043, 2.7043)	0.000000

Table A.4.43: *BS-SSM$_{\beta}$ using an $m = 100$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0364 (0.0364, 0.0364)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1253)	0.017535
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.5190 (-0.5190, -0.5190)	0.000000
$\widehat{\beta}_{\sigma}$	2.7063 (2.7063, 2.7063)	0.000000

Table A.4.44: *BS-SSM$_{\beta}$ using an $m = 200$ point grid and truncation $b_{\max} = 1.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9993 (0.9993, 0.9993)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0001 (0.0001, 0.0001)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.4009 (0.4009, 0.4009)	0.000000
$\widehat{\mu}_{\text{tot}}$	0.4349 (—, —)	—
$\widehat{\sigma}$	0.1353 (0.0179, 0.2527)	0.059892
$\widehat{\beta}_{\mu}$	-0.0659 (-0.0659, -0.0659)	0.000000
$\widehat{\beta}_{\sigma}$	3.6550 (-5.0218, 12.3317)	4.426908

Table A.4.45: *BS-SSM$_{\beta}$ using an $m = 20$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0705 (0.0705, 0.0705)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017529
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.2683 (-0.2683, -0.2683)	0.000000
$\widehat{\beta}_{\sigma}$	1.3990 (1.3990, 1.3990)	0.000000

Table A.4.46: *BS-SSM$_{\beta}$ using an $m = 70$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0705 (0.0705, 0.0705)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017531
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004942
$\widehat{\beta}_{\mu}$	-0.2683 (-0.2683, -0.2683)	0.000000
$\widehat{\beta}_{\sigma}$	1.3993 (1.3993, 1.3993)	0.000000

Table A.4.47: *BS-SSM$_{\beta}$ using an $m = 100$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0704 (0.0704, 0.0704)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017530
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004942
$\widehat{\beta}_{\mu}$	-0.2684 (-0.2684, -0.2684)	0.000000
$\widehat{\beta}_{\sigma}$	1.3999 (1.3999, 1.3999)	0.000000

Table A.4.48: *BS-SSM$_{\beta}$ using an $m = 200$ point grid and truncation $b_{\max} = 2.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9981 (0.9981, 0.9981)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0002 (0.0002, 0.0002)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.1012 (0.1012, 0.1012)	0.000000
$\widehat{\mu}_{\text{tot}}$	0.1353 (—, —)	—
$\widehat{\sigma}$	0.1140 (-0.0001, 0.2282)	0.058230
$\widehat{\beta}_{\mu}$	-0.0297 (-0.0297, -0.0297)	0.000000
$\widehat{\beta}_{\sigma}$	2.0544 (-4.6174, 8.7262)	3.403985

Table A.4.49: *BS-SSM$_{\beta}$ using an $m = 20$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9995 (0.9995, 0.9995)	0.000001
$\widehat{\sigma}_{\varepsilon}$	0.0001 (0.0001, 0.0001)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	1.1720 (0.8863, 1.4577)	0.145788
$\widehat{\mu}_{\text{tot}}$	1.2060 (—, —)	—
$\widehat{\sigma}$	1.0067 (1.0067, 1.0067)	0.000000
$\widehat{\beta}_{\mu}$	-0.8772 (-3.2273, 1.4729)	1.199027
$\widehat{\beta}_{\sigma}$	-0.4718 (-0.4718, -0.4718)	0.000000

Table A.4.50: *BS-SSM$_{\beta}$ using an $m = 40$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9986 (0.9986, 0.9986)	0.000003
$\widehat{\sigma}_{\varepsilon}$	0.0001 (0.0001, 0.0001)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.4133 (0.4133, 0.4133)	0.000000
$\widehat{\mu}_{\text{tot}}$	0.4474 (—, —)	—
$\widehat{\sigma}$	0.1283 (-0.1555, 0.4122)	0.144821
$\widehat{\beta}_{\mu}$	-1.7006 (-1.7006, -1.7006)	0.000000
$\widehat{\beta}_{\sigma}$	10.5171 (-41.0901, 62.1244)	26.330239

Table A.4.51: *BS-SSM$_{\beta}$ using an $m = 70$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0769 (0.0769, 0.0769)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017532
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.2459 (-0.2459, -0.2459)	0.000000
$\widehat{\beta}_{\sigma}$	1.2824 (1.2824, 1.2824)	0.000000

Table A.4.52: *BS-SSM$_{\beta}$ using an $m = 100$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0769 (0.0769, 0.0769)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017531
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.2459 (-0.2459, -0.2459)	0.000000
$\widehat{\beta}_{\sigma}$	1.2825 (1.2825, 1.2825)	0.000000

Table A.4.53: *BS-SSM$_{\beta}$ using an $m = 200$ point grid and truncation $b_{\max} = 3.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000001
$\widehat{\sigma}_{\varepsilon}$	0.0000 (0.0000, 0.0000)	0.000002
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	-700.6359 (-702.2372, -699.0346)	0.816979
$\widehat{\mu}_{\text{tot}}$	-700.6019 (—, —)	—
$\widehat{\sigma}$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{\beta}_{\mu}$	-3508.6879 (-3516.6923, -3500.6835)	4.083880
$\widehat{\beta}_{\sigma}$	-3527.8868 (-3527.8868, -3527.8868)	0.000000

Table A.4.54: *BS-SSM$_{\beta}$ using an $m = 20$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.2385 (-0.9011, 1.3781)	0.581422
$\widehat{\mu}_{\text{tot}}$	0.2725 (—, —)	—
$\widehat{\sigma}$	1.9752 (1.9752, 1.9752)	0.000000
$\widehat{\beta}_{\mu}$	1.6992 (-9.3980, 12.7963)	5.661823
$\widehat{\beta}_{\sigma}$	7.5654 (7.5654, 7.5654)	0.000000

Table A.4.55: *BS-SSM$_{\beta}$ using an $m = 40$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	1.0000 (1.0000, 1.0000)	0.000000
$\widehat{\sigma}_{\varepsilon}$	0.0000 (0.0000, 0.0000)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0064 (-7.1635, 7.1764)	3.658142
$\widehat{\mu}_{\text{tot}}$	0.0405 (—, —)	—
$\widehat{\sigma}$	0.0035 (0.0035, 0.0035)	0.000000
$\widehat{\beta}_{\mu}$	-0.8171 (-2.9652, 1.3311)	1.095999
$\widehat{\beta}_{\sigma}$	1.0604 (1.0604, 1.0604)	0.000000

Table A.4.56: *BS-SSM$_{\beta}$ using an $m = 70$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0769 (0.0769, 0.0769)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017532
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.2459 (-0.2459, -0.2459)	0.000000
$\widehat{\beta}_{\sigma}$	1.2824 (1.2824, 1.2824)	0.000000

Table A.4.57: *BS-SSM$_{\beta}$ using an $m = 100$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*

Parameter	Estimate (95% CI)	Std. Error
$\widehat{\rho}$	0.9827 (0.9794, 0.9861)	0.001701
$\widehat{\sigma}_{\varepsilon}$	0.0769 (0.0769, 0.0769)	0.000000
$\widehat{q}$	0.0340 (0.0322, 0.0359)	0.000922
$\widehat{\mu}_{\text{cap}}$	0.0909 (0.0565, 0.1252)	0.017531
$\widehat{\mu}_{\text{tot}}$	0.1249 (—, —)	—
$\widehat{\sigma}$	0.1306 (0.1209, 0.1403)	0.004941
$\widehat{\beta}_{\mu}$	-0.2459 (-0.2459, -0.2459)	0.000000
$\widehat{\beta}_{\sigma}$	1.2824 (1.2824, 1.2824)	0.000000

Table A.4.58: *BS-SSM$_{\beta}$ using an $m = 200$ point grid and truncation $b_{\max} = 4.0$. Parameter estimates with 95% confidence intervals in parentheses.*