Efficient Estimation of Stochastic Parameters: A GLS Approach

Da Huo, Da

MIT Sloan School of Management

15 January 2024

Efficient Estimation of Stochastic Parameters: A GLS Approach

by

Da Huo

B.E. Finance, Nankai University, 2022

Submitted to the MIT Sloan School of Management
in partial fulfillment of the requirements for the degree of

MASTER OF FINANCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by: Da Huo
MIT Sloan School of Management
January 15, 2024

Certified by: Hui Chen
Nomura Professor of Finance, Thesis Supervisor

Accepted by: Urmi Samadar
Assistant Dean, MIT Sloan Master of Finance Program
MIT Sloan School of Management

Efficient Estimation of Stochastic Parameters: A GLS Approach

by

Da Huo

Submitted to the MIT Sloan School of Management
on January 15, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF FINANCE

ABSTRACT

This thesis presents a novel rolling GLS-based model to improve the precision of time-varying parameter estimates in dynamic linear models. Through rigorous simulations, the rolling GLS model exhibits enhanced accuracy in scenarios with smaller sample sizes and maintains its efficacy when the normality assumption is relaxed, distinguishing it from traditional models like Kalman Filters. Furthermore, the thesis expands on the model to tackle more complex stochastic structures and validates its effectiveness through practical applications to real-world financial data, like inflation risk premium estimations. The research culminates in offering a robust tool for financial econometrics, enhancing the reliability of financial analyses and predictions.

Thesis supervisor: Hui Chen
Title: Nomura Professor of Finance

Acknowledgments

I would like to express my deepest gratitude to my thesis supervisor, Prof. Hui Chen, whose expertise and insight have been instrumental in guiding me through the complexities of my research. His patience and scholarly guidance have been invaluable, not only for this thesis but as a foundation for my future endeavors in the field of finance.

I am also indebted to my friend and fellow thesis writer, Tyson Dennis-Sharma, for the enlightening discussions and for challenging me with new ideas. His perspective and intellectual camaraderie have significantly enriched my research experience.

I extend my thanks to my peers and mentors from MIT and Nankai University, whose rigorous debates and shared wisdom have contributed to my academic growth. A special mention goes to Prof. Anna Mikusheva, for her support and constructive criticisms during the development of this thesis.

My heartfelt appreciation also goes to the administrative and academic staff of MIT Sloan School of Management, who have provided an environment conducive to research and learning.

Finally, I would like to thank my family for their unwavering support and encouragement. Their belief in my abilities has been a constant source of strength and motivation.

This journey would not have been possible without the collective support and encouragement of each one mentioned above and many others who have contributed in various capacities.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

The aim of this thesis is to address a significant challenge in financial econometrics: the efficient estimation of stochastic parameters, particularly time-varying parameters. The focus is on employing a rolling Generalized Least Squares (GLS) type GMM model to minimize the Root Mean Square Error (RMSE) from the estimation of these parameters.

The model setup involves assuming a linear state model, and a linear relationship between the measurement and the underlying state:

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t \tag{1.1}$$

$$\boldsymbol{\beta}_t = \sum_{j=1}^{p} A_j \boldsymbol{\beta}_{t-j} + \boldsymbol{\nu}_t \tag{1.2}$$

A rolling window regression is used for estimation. Here, the "error" term is defined with respect to the beta of interest. Normally, we would like to recover the average beta for the full time series. However, here, we are interested in the beta for each time stamp individually. In a causal model, this will be the last beta in the window. By contrast, in a non-causal model, this will be the middle beta in the window.

GLS is an advanced approach that accounts for potential heteroskedasticity or autocorrelation in the error terms. The GLS method aims to provide more efficient and unbiased estimates compared to OLS, especially in the presence of non-constant variance in the error terms. This is ideal for my application here, as error terms from different lags are indeed correlated with each other.

This gives us the GLS estimator of

$$\widehat{\boldsymbol{\beta}}_t = (X_{t-k,t}^\top \Omega_{t-k,t}^{-1} X_{t-k,t})^{-1} (X_{t-k,t}^\top \Omega_{t-k,t}^{-1} \boldsymbol{y}_{t-k,t}) \tag{1.3}$$

, where k is the maximum look-back period.

The primary challenge tackled in this thesis is deriving

$$\Omega_{t-k,t} = \mathrm{Var}(u_{t-k,t} | X_{t-k,t}, \boldsymbol{\beta}_t) \tag{1.4}$$

, where u_{t-j} is the estimation error

$$u_{t-j} = y_{t-j} - \boldsymbol{x}_{t-j}^{\top}\boldsymbol{\beta}_t \tag{1.5}$$

As will be shown in the subsequent chapters, the GLS model presented in this study demonstrates significant improvements, particularly in handling smaller sample sizes and offering robust performance when normality assumptions are compromised, setting a new benchmark in financial modeling.

Following this overview, the thesis unfolds over five additional chapters and two sections. The rest of Chapter 1 summarizes the motivation for this thesis, as well as various methodologies previously applied in this field. In Chapter 2, I establish the foundational models and estimators crucial for the subsequent empirical methodology. Chapter 3 details the feasible application of the theoretical models, including the Method-of-Moments estimator and the iterative estimator. Chapter 4 presents a series of simulations conducted to test the robustness and efficacy of the proposed models under known and unknown parameters. In Chapter 5, I apply the models to actual financial data to illustrate their practical utility in portfolio construction, factor estimation, and inflation risk premium analysis. Finally, Chapter 6 summarizes the findings and discusses potential avenues for future research.

1.2 Motivation

The motivation for this study stems from the need to enhance the precision and reliability of financial models, especially in the context of rapidly changing market dynamics. Current methods, while widely used, have limitations in handling the time-varying nature of modern financial relationships. The proposed rolling GLS filter aims to address these shortcomings by providing a more robust and dynamic framework for parameter estimation.

1.2.1 Factor Risk Premium

In the process of factor risk premium estimation, we observe changes in the characteristics and factor loadings of firms over time. When a firm increases its leverage, its beta rises, implying a higher sensitivity to market movements. Similarly, as a firm grows in size, its loading on the size factor decreases, illustrating a shift in its risk profile relative to the market.

Beyond these traditional metrics, the thesis also explores other factors that are less frequently accounted for in standard models, such as the inflation factor. These factors represent the unique aspects of a firm's financial behavior that are not captured by conventional risk factors.

1.2.2 Portfolio Construction

The construction of a portfolio requires efficient estimation of expected returns. This step is critical for achieving an optimal allocation of assets that maximizes returns while minimizing risk. The thesis delves into the methodologies used for estimating these expected returns, particularly focusing on the application and implications of my proposed filter in this context.

1.2.3 Portfolio Evaluation

Evaluating a portfolio involves assessing the changing nature of its alpha, or excess return, over time. It is observed that the alpha of a managed portfolio may vary, often exhibiting low market beta during normal times and high market beta during crises. This phenomenon highlights the need for dynamic models that can adapt to changing market conditions and more accurately capture the risk-return profile of the portfolio across different market phases.

1.3 Existing Approaches

1.3.1 Equal-weighted Rolling Ordinary Least Square

One of the foundational methods for estimating time-varying parameters is the equal-weighted rolling OLS. In practice, empirical approaches often rely on this method to estimate time-varying parameters. It is given by:

$$\widehat{\beta}_t = \left(\sum_{i=0}^{k} x_{t-i} x_{t-i}' \right)^{-1} \sum_{i=0}^{k} x_{t-i} y_{t-i}, \tag{1.6}$$

where $x_{t-i},\ y_{t-i}$ denotes the observed independent and dependent variable at time $t-i$, and k is the number of periods in the rolling window. The equal-weighted rolling average assumes that all observations within the window contribute equally to the estimation of the current parameter value, β_t. This method is particularly useful for its simplicity and ease of computation.

However, it may not be ideal when recent observations are more indicative of the current state, a scenario where weighted schemes could provide a more accurate reflection of the parameter dynamics. What's worse, the selected window horizon for estimation is often based on the performance optimization of certain asset pricing models. Parameters estimated in this manner can introduce significant biases. This ad hoc approach can lead to spurious conclusions or "cherry-picking" results that do not accurately reflect the underlying financial dynamics.

1.3.2 Linearly Decaying Rolling Weighted Least Square

The linearly decaying WLS refines the concept of the rolling OLS by assigning weights that linearly decrease for observations further in the past. This method can be mathematically represented as:

$$\widehat{\beta}_t = \left(\sum_{i=0}^{k} (k-i+1) x_{t-i} x_{t-i}' \right)^{-1} \sum_{i=0}^{k} (k-i+1) x_{t-i} y_{t-i}, \tag{1.7}$$

where each observation is weighted by its lag index $k-i+1$, ensuring that more recent data points have a larger influence on the estimate.

This filter is motivated by the Bartlett kernel adopted in the Newey and West 1987, though we have a completely different problem here. It is more aligned with the belief that recent observations may be more relevant in representing the current state.

However, it's still not ideal. One reason being that the shape of the kernel is not well justified. One would expect a better performance by moving towards non-linear kernels, such as the Parzen Window, Tukey-Hanning Window and the Quadratic Spectral Kernel. More importantly, we would still need to hand-pick the size of the estimation window, which again leads to suspicious data mining.

1.3.3 Kalman Filter

The Kalman Filter is a recursive solution to the linear Bayesian filtering problem and is widely used for estimating time-varying parameters in systems governed by linear stochastic difference equations (Kalman 1960).

This is probably the most popular model adopted in these "Dynamic Linear Model" set-ups. It is based on maximum-likelihood algorithms that evaluate the probability density function (pdf) of the hidden state recursively in a Bayesian manner.

Its formulation is as follows:

$$\widehat{\boldsymbol{\beta}}_{t|t-1} = A\widehat{\boldsymbol{\beta}}_{t-1|t-1}, \tag{1.8}$$

$$P_{t|t-1} = AP_{t-1|t-1}A^\top + Q, \tag{1.9}$$

$$K_t = P_{t|t-1}H^\top (HP_{t|t-1}H^\top + R)^{-1}, \tag{1.10}$$

$$\widehat{\boldsymbol{\beta}}_{t|t} = \widehat{\boldsymbol{\beta}}_{t|t-1} + K_t(y_t - H\widehat{\boldsymbol{\beta}}_{t|t-1}), \tag{1.11}$$

$$P_{t|t} = (I - K_t H)P_{t|t-1}. \tag{1.12}$$

Here, A represents the state transition matrix, Q the covariance matrix of the process noise, H the observation matrix, R the covariance matrix of the observation noise, K_t the Kalman gain, and P the covariance matrix of the estimated parameters.

The Kalman Filter's strength lies in its optimality and efficiency. However, Kalman filtering imposes several restrictive assumptions. First and foremost, in the "Bayesian Updating" step, the Kalman filter assumes that errors are Gaussian, which is not widely true in practice. In addition, the likelihood function may be poorly defined in non-stationary cases, leading to non-vanishing priors. Addressing these challenges, the GLS model presented in this study demonstrates significant improvements.

1.3.4 Kalman Smoother

The Kalman Smoother extends the Kalman Filter to provide smoothed estimates of past states by incorporating all available observations. The key equations for the Kalman Smoother are:

$$\widehat{\boldsymbol{\beta}}_{t|T} = \widehat{\boldsymbol{\beta}}_{t|t} + L_t(\widehat{\boldsymbol{\beta}}_{t+1|T} - A\widehat{\boldsymbol{\beta}}_{t|t}), \tag{1.13}$$

$$P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^\top, \tag{1.14}$$

$$L_t = P_{t|t}A^\top P_{t+1|t}^{-1}. \tag{1.15}$$

Where T denotes the total number of observations, and L_t is the smoother gain. This technique is particularly powerful for signal extraction and smoothing noisy data series. A foundational reference for understanding the Kalman Smoother is the work by Rauch, Tung, and Striebel 1965 on state estimation for linear systems subject to Gaussian noise, commonly referred to as the RTS Smoother.

1.3.5 Hamilton Filter

The Hamilton Filter, is a pivotal tool in econometrics for analyzing nonstationary time series with regime changes (Hamilton 1989). It is particularly effective in capturing the dynamics of economic time series that exhibit phases such as growth and recession. The filter operates within the framework of a Markov-switching model and can be represented as:

$$y_t = \beta_0 + \beta_1 S_t + \varepsilon_t \tag{1.16}$$

$$S \in \{0, 1\} \tag{1.17}$$

$$P(S_t = 1|S_{t-1} = 0) = 1 - q \tag{1.18}$$

$$P(S_t = 1|S_{t-1} = 1) = p \tag{1.19}$$

where S_t is the state vector, $\boldsymbol{y}_t$ the observed time series, and $\boldsymbol{\beta}$ the set of model parameters.

1.3.6 Particle Filters

Particle Filters, also known as Sequential Monte Carlo methods, are crucial in state estimation for non-linear and non-Gaussian models. They utilize a set of random samples, or particles, to approximate the posterior distributions of state variables. The weights of these particles are updated based on their likelihood:

$$\widehat{\boldsymbol{x}}_t = \sum_{i=1}^{N} w_t^{(i)} \boldsymbol{x}_t^{(i)}, \tag{1.20}$$

where $\boldsymbol{x}_t^{(i)}$ are the particles and $w_t^{(i)}$ their weights. For further details, see Doucet and Johansen 2011 on particle filtering and smoothing.

1.3.7 Extended Kalman Filter

The Extended Kalman Filter (EKF) is designed for systems with non-linear dynamics. It linearizes about the current estimate to handle non-linearities in the state and observation models:

$$\widehat{\boldsymbol{x}}_{t|t-1} = f(\widehat{\boldsymbol{x}}_{t-1|t-1}, \boldsymbol{u}_t), \tag{1.21}$$
$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^\top + Q_t, \tag{1.22}$$

with $f(\cdot)$ as the non-linear state transition function, and F_t its Jacobian. Julier and Uhlmann 1997 provides an extensive discussion on EKF.

1.3.8 Unscented Kalman Filter

The Unscented Kalman Filter (UKF) addresses the shortcomings of EKF for highly non-linear systems using a deterministic sampling approach:

$$\boldsymbol{\chi}_t = \text{GenerateSigmaPoints}(\widehat{\boldsymbol{x}}_{t-1}, P_{t-1}), \tag{1.23}$$
$$\widehat{\boldsymbol{x}}_{t|t-1} = \sum_i w_m^{(i)} f(\boldsymbol{\chi}_t^{(i)}), \tag{1.24}$$
$$P_{t|t-1} = \sum_i w_c^{(i)} [\boldsymbol{\chi}_t^{(i)} - \widehat{\boldsymbol{x}}_{t|t-1}][\boldsymbol{\chi}_t^{(i)} - \widehat{\boldsymbol{x}}_{t|t-1}]^\top + Q_t, \tag{1.25}$$

UKF employs sigma points to approximate the mean and covariance of the state distribution. Wan and Van Der Merwe 2000 offers a comprehensive explanation of UKF.

1.3.9 Hodrick-Prescott Filter

The Hodrick-Prescott (HP) Filter is a widely-used tool for extracting the cyclical component of a time series from raw data (Hodrick and Prescott 1997). It is formulated as the solution to the following optimization problem:

$$\min_{\{\mu_t\}} \left\{ \sum_{t=1}^{T} (y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} [(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2 \right\}, \tag{1.26}$$

where y_t is the observed time series, μ_t the trend component, and λ the smoothing parameter. The HP filter differentiates between short-term fluctuations and long-term trends, making it useful in macroeconomic analysis.

However, it should be treated with caution. HP Filter averages over today's, yesterday's and tomorrow's data. If we are interested in whether one series leads (or perhaps causes) another, we may mess up the relationship by filtering. In addition, HP can generate spurious cycles. Cogley and Nason 1995 generated random-walk data, applied HP, and found cycles. Baxter-King Filter might also generate spurious cycles. In general, we know the theory of these filters, but we don't know their stochastic properties so well.

1.3.10 Baxter-King Filter

The Baxter-King Filter is another popular method used for extracting cyclical components from time series data, especially in macroeconomic analysis (Baxter and King 1999). Unlike

the Hodrick-Prescott Filter, the Baxter-King filter aims to approximate an ideal band-pass filter that passes frequencies contained in a pre-specified band while attenuating frequencies outside this band.

The mathematical formulation of the Baxter-King filter involves applying a moving average filter to the time series data. This filter is defined as follows:

$$B(L) = \sum_{j=-q}^{q} b_j L^j \tag{1.27}$$

where $B(L)$ is the filter in terms of the lag operator L, and b_j are the filter coefficients designed to target specific cyclical frequencies.

The spectral density of the filtered series y_t is related to the spectral density of x_t through the relationship:

$$S_y(\omega) = S_x(\omega)|B(e^{i\omega})|^2 \tag{1.28}$$

where $S_y(\omega)$ and $S_x(\omega)$ are the spectral densities of y_t and x_t, respectively, and ω represents the frequency.

The key insight of the Baxter-King filter lies in its ability to provide an approximation to the ideal filter that can be computed with finite data. This is particularly useful in empirical analysis where the ideal filter would require an infinite amount of data. The Baxter-King filter achieves this by truncating the infinite sum in the ideal filter to a finite number of terms, determined by the chosen number of lags K, and adjusting the coefficients to satisfy the condition $B(1) = 0$, which removes the deterministic trend from the series.

The filter coefficients b_j are selected to minimize the difference between the actual frequency response of the filter and the desired frequency response over the target range of frequencies. This leads to the truncated filter that approximates the ideal response as closely as possible given the constraints of finite data.

The effectiveness of the Baxter-King filter can be evaluated by comparing the gain of the filter, $|B(e^{i\omega})|^2$, against the desired frequency response. This comparison demonstrates the filter's ability to isolate the cyclical components associated with the business cycle from other components, such as the trend and seasonal effects, in economic time series data.

Chapter 2

Theoretical Derivation

This chapter is composed of several key sections, each delving into different aspects of the theoretical framework underpinning this study. The chapter begins with an exploration of the 'Simple Stochastic Mean Model,' where the model's set-up and estimator formulations, both causal and non-causal, are discussed. This is followed by a similar analysis for the 'General Stochastic Mean Model.' The chapter then progresses to a detailed examination of the 'VAR(1) Stochastic Beta Model,' outlining its foundational principles and estimator strategies. Finally, the 'General Stochastic Beta Model' is presented, completing the theoretical exploration.

2.1 Simple Stochastic Mean Model

2.1.1 Model Set-up

The simple stochastic mean model serves as a foundational framework for the exploration of time-varying parameter estimations. The model encapsulates the core dynamics that will be extended in more complex formulations discussed later in this thesis. It is defined by the following equations:

$$y_t = \mu_t + \varepsilon_t, \tag{2.1}$$
$$\mu_t = \mu_{t-1} + \nu_t, \tag{2.2}$$

, where y_t denotes the observed variable at time t, capturing the realizations of the stochastic process. The term μ_t represents the stochastic mean, which is the principal time-varying parameter of interest, highlighting the mean's evolution over time. The noise component ε_t accounts for random fluctuations around the mean, while ν_t signifies the incremental stochastic shocks that drive the mean's progression.

The stochastic properties of the noise and shock components are characterized by uncorrelated white noises, as specified below:

$$\varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2), \tag{2.3}$$
$$\nu_t \sim \mathcal{WN}(0, \sigma_\nu^2), \tag{2.4}$$
$$\text{Cov}(\varepsilon_t, \nu_s) = 0 \text{ for all } s, t, \tag{2.5}$$

, which ensures that the noise and the shocks are uncorrelated across time, maintaining the integrity of the mean as a purely stochastic element devoid of systematic influence from the noise.

Given these assumptions, the primary objective of my analysis is the robust recovery of the time-varying parameter μ_t from the observable series y_t. This task lays the groundwork for the subsequent development of estimation techniques that can adeptly handle the intrinsic variability characteristic of financial time series.

2.1.2 Causal Estimator Formulation

As is revealed from the overview, my GLS estimator for μ_t will be in the form of

$$\widehat{\mu}_t = (\iota^\top \Omega^{-1} \iota)^{-1} (\iota^\top \Omega^{-1} \boldsymbol{y}_{t-k,t}) \tag{2.6}$$

The development of the GLS estimator for the stochastic mean μ_t is predicated on the precise articulation of the variance of the estimation error, Ω. This variance is a cornerstone in the implementation of the GLS methodology, as it encapsulates the error dynamics essential for obtaining efficient estimates.

To elucidate the estimation of μ_t, I consider the observable y's as functions of μ_t, leading to a system of equations that defines the relationship between observed values and the stochastic mean, coupled with the noise and underlying shocks:

$$y_t = \mu_t + \varepsilon_t \tag{2.7}$$
$$y_{t-1} = \mu_{t-1} + \varepsilon_{t-1} = \mu_t - \nu_t + \varepsilon_{t-1} \tag{2.8}$$
$$y_{t-2} = \mu_t - \nu_t - \nu_{t-1} + \varepsilon_{t-2} \tag{2.9}$$
$$\cdots\cdots \tag{2.10}$$
$$y_{t-k} = \mu_t - \nu_t - \cdots - \nu_{t-k+1} + \varepsilon_{t-k} \tag{2.11}$$

Subtracting μ_t from each observable yields the estimation errors, which are essentially the discrepancies between the observed values and the stochastic mean:

$$u_t = y_t - \mu_t = \varepsilon_t \tag{2.12}$$
$$u_{t-1} = y_{t-1} - \mu_t = -\nu_t + \varepsilon_{t-1} \tag{2.13}$$
$$u_{t-2} = y_{t-2} - \mu_t = -\nu_t - \nu_{t-1} + \varepsilon_{t-2} \tag{2.14}$$
$$\cdots\cdots \tag{2.15}$$
$$u_{t-k} = y_{t-k} - \mu_t = -\nu_t - \cdots - \nu_{t-k+1} + \varepsilon_{t-k} \tag{2.16}$$

The core of our estimator lies in the conditional variance-covariance matrix $\Omega_{\boldsymbol{u}|\mu_t}$, which is structured as a block matrix to reflect the variances and covariances of the estimation errors $\begin{bmatrix} u_t & u_{t-1} & \cdots & u_{t-k} \end{bmatrix}$:

$$\Omega_{\boldsymbol{u}|\mu_t} = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 + \sigma_\nu^2 & \sigma_\nu^2 & \sigma_\nu^2 & \cdots & \sigma_\nu^2 \\ 0 & \sigma_\nu^2 & \sigma_\varepsilon^2 + 2\sigma_\nu^2 & 2\sigma_\nu^2 & \cdots & 2\sigma_\nu^2 \\ 0 & \sigma_\nu^2 & 2\sigma_\nu^2 & \sigma_\varepsilon^2 + 3\sigma_\nu^2 & \cdots & 3\sigma_\nu^2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \sigma_\nu^2 & 2\sigma_\nu^2 & 3\sigma_\nu^2 & \cdots & \sigma_\varepsilon^2 + k\sigma_\nu^2 \end{bmatrix} \tag{2.17}$$

Finally, the GLS estimator $\widehat{\mu}_t^{GLS}$ is obtained by applying the inverse of this matrix to the vector of observed values, facilitating a Best Linear Unbiased Estimation (BLUE) of the stochastic mean:

$$\widehat{\mu}_t^{GLS} = (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \iota)^{-1} (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \boldsymbol{y}_{t-k,t}). \tag{2.18}$$

This formulation of the GLS estimator not only underscores the importance of accounting for the variance-covariance structure in estimation, but also highlights the estimator's capacity to harness this information for more precise inference in the presence of time-varying parameters. Since the estimator is formulated under the GLS framework, the standard GLS standard error can be applied to construct confidence intervals and perform hypothesis testing.

2.1.3 Non-Causal Estimator Formulation

Another innovative aspect of this research involves the formulation of a non-causal estimator that leverages both past and future information to estimate the stochastic mean μ_t. This approach is also grounded in the derivation of Ω, the variance of the estimation error, which is a critical component of the GLS estimator's framework.

To establish the foundation for defining the estimation errors, we express all observable variables $y_{t-k:t+k}$ as functions of the stochastic mean μ_t. The expressions are as follows:

$$y_{t-k} = \mu_t - \nu_t - \cdots - \nu_{t-k+1} + \varepsilon_{t-k} \tag{2.19}$$
$$y_{t-1} = \mu_{t-1} + \varepsilon_{t-1} = \mu_t - \nu_t + \varepsilon_{t-1} \tag{2.20}$$
$$y_t = \mu_t + \varepsilon_t \tag{2.21}$$
$$y_{t+1} = \mu_{t+1} + \varepsilon_{t+1} = \mu_t + \nu_{t+1} + \varepsilon_{t+1} \tag{2.22}$$
$$y_{t+k} = \mu_t + \nu_{t+1} + \cdots + \nu_{t+k} + \varepsilon_{t+k} \tag{2.23}$$

Subsequently, the estimation errors can be articulated as:

$$u_{t-k} = y_{t-k} - \mu_t = -\nu_t - \cdots - \nu_{t-k+1} + \varepsilon_{t-k} \tag{2.24}$$
$$u_{t-1} = y_{t-1} - \mu_t = -\nu_t + \varepsilon_{t-1} \tag{2.25}$$
$$u_t = y_t - \mu_t = \varepsilon_t \tag{2.26}$$
$$u_{t+1} = y_{t+1} - \mu_t = \nu_{t+1} + \varepsilon_{t+1} \tag{2.27}$$
$$u_{t+k} = y_{t+k} - \mu_t = \nu_{t+1} + \cdots + \nu_{t+k} + \varepsilon_{t+k} \tag{2.28}$$

The construction of the conditional variance-covariance matrix of the estimation error vector is intricate, as it involves a block diagonal arrangement:

$$\Omega_{\boldsymbol{u}|\mu_t} = \text{blkdiag}(\{\Omega_{sub}^{\top'}, \sigma_\varepsilon^2, \Omega_{sub}\}), \tag{2.29}$$

where the submatrix Ω_{sub} is defined to capture the variances and covariances that evolve

over different time horizons:

$$\Omega_{sub} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\nu^2 & \sigma_\nu^2 & \sigma_\nu^2 & \cdots & \sigma_\nu^2 & \\ \sigma_\nu^2 & \sigma_\varepsilon^2 + 2\sigma_\nu^2 & 2\sigma_\nu^2 & \cdots & 2\sigma_\nu^2 & \\ \sigma_\nu^2 & 2\sigma_\nu^2 & \sigma_\varepsilon^2 + 3\sigma_\nu^2 & \cdots & 3\sigma_\nu^2 & \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \sigma_\nu^2 & 2\sigma_\nu^2 & 3\sigma_\nu^2 & \cdots & \sigma_\varepsilon^2 + k\sigma_\nu^2 & \end{bmatrix} \tag{2.30}$$

This matrix's unique structure, with variances that increase with the temporal distance from the mean estimation period and constant covariances, is a reflection of the underlying stochastic processes.

To ensure symmetry within the non-causal framework, consider the transpose of Ω_{sub} with respect to the anti-diagonal, noted as $\Omega_{sub}^{\top'}$. This transformation is a non-standard operation in matrix algebra but is indispensable for the non-causal estimator's architecture.

The culmination of these derivations is, again, the GLS estimator, $\widehat{\mu}_t^{GLS}$, formulated as:

$$\widehat{\mu}_t^{GLS} = (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \iota)^{-1} (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \boldsymbol{y}_{t-k,t+k}). \tag{2.31}$$

This estimator stands as a testament to the power of GLS methodology, offering an efficient and unbiased estimation of μ_t by adeptly using the structured error variances encapsulated within the earlier derivations.

2.2 General Stochastic Mean Model

2.2.1 Model Set-up

In this model, we still observe y_t, a variable comprising the stochastic mean μ_t and a noise component ε_t. In addition, the noise component is uncorrelated with the stochastic mean at all times.

However, as a general case, the stochastic mean can be any stationary time series, characterized by an auto-covariance function γ_k. This is formalized as:

$$y_t = \mu_t + \varepsilon_t, \tag{2.32}$$

$$\mu_t \text{ is stationary with auto-covariance function } \gamma_k, \tag{2.33}$$

$$\text{Cov}(\varepsilon_t, \mu_s) = \mathbf{0} \text{ for all } s, t. \tag{2.34}$$

2.2.2 Causal Estimator Formulation

The causal estimator for μ_t still relies on the Generalized Least Squares (GLS) method. The detailed derivation of the GLS estimator is beyond the scope of this section, but interested readers are encouraged to contact the author for further insights.

The GLS estimator is expressed as:

$$\widehat{\mu}_t^{GLS} = (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \iota)^{-1} (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \boldsymbol{y}_{t-k,t}), \tag{2.35}$$

where $\Omega_{\boldsymbol{u}|\mu_t}$ encapsulates the variances and covariances of the estimation errors, integrating the stationary properties of μ_t:

$$\Omega_{\boldsymbol{u}|\mu_t} = \sigma_\varepsilon^2 \cdot I_{k\times k} + \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 2\gamma_0 - 2\gamma_1 & \gamma_0 - \gamma_2 & \cdots & \gamma_0 - \gamma_1 + \gamma_{k-1} - \gamma_k \\ 0 & \gamma_0 - \gamma_2 & 2\gamma_0 - 2\gamma_2 & \cdots & \gamma_0 - \gamma_2 + \gamma_{k-2} - \gamma_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \gamma_0 - \gamma_1 + \gamma_{k-1} - \gamma_k & \gamma_0 - \gamma_2 + \gamma_{k-2} - \gamma_k & \cdots & 2\gamma_0 - 2\gamma_k \end{bmatrix}. \tag{2.36}$$

This GLS estimator takes into account the stationarity of the stochastic mean, μ_t, and the orthogonality of the noise term with respect to μ_t, thus enabling a precise recovery of the underlying stochastic process driving the observed y_t.

2.2.3 Non-Causal Estimator Formulation

Similar to Section 2.1.3, due to the inherent symmetry in time series, the non-causal two-sided GLS estimator is expressed as:

$$\widehat{\mu}_t^{GLS} = (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \iota)^{-1} (\iota^\top \Omega_{\boldsymbol{u}|\mu_t}^{-1} \boldsymbol{y}_{t-k,t+k}), \tag{2.37}$$

where

$$\Omega_{\boldsymbol{u}|\mu_t} = \sigma_\varepsilon^2 \cdot I_{(2k+1)\times(2k+1)} + \text{blkdiag}(\{\Omega_{sub}^{\top'}, 0, \Omega_{sub}\}), \tag{2.38}$$

and

$$\Omega_{sub} = \begin{bmatrix} 2\gamma_0 - 2\gamma_1 & \gamma_0 - \gamma_2 & \cdots & \gamma_0 - \gamma_1 + \gamma_{k-1} - \gamma_k \\ \gamma_0 - \gamma_2 & 2\gamma_0 - 2\gamma_2 & \cdots & \gamma_0 - \gamma_2 + \gamma_{k-2} - \gamma_k \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_0 - \gamma_1 + \gamma_{k-1} - \gamma_k & \gamma_0 - \gamma_2 + \gamma_{k-2} - \gamma_k & \cdots & 2\gamma_0 - 2\gamma_k \end{bmatrix}. \tag{2.39}$$

2.3 VAR(1) Stochastic Beta Model

2.3.1 Model Set-up

In the Vector Autoregressive (VAR) Stochastic Beta Model, I model the observed variable y_t as a product of a time-varying parameter vector $\boldsymbol{\beta}_t$ and a vector of regressors $\boldsymbol{x}_t$, with an additive noise component ε_t with potential heteroskedasticity and auto-correlation:

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t, \tag{2.40}$$

$$\text{where } \mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}] = \boldsymbol{0}, \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\boldsymbol{X}] = \Sigma_\varepsilon. \tag{2.41}$$

The time-varying parameter $\boldsymbol{\beta}_t$ is modeled to follow a process centered around a mean vector $\bar{\boldsymbol{\beta}}$, with innovations $\boldsymbol{\nu}_t$:

$$\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}} = A(\boldsymbol{\beta}_{t-1} - \bar{\boldsymbol{\beta}}) + \boldsymbol{\nu}_t, \tag{2.42}$$

$$\text{where } \boldsymbol{\nu}_t \sim \mathcal{WN}(\boldsymbol{0}, \Sigma_\nu), \tag{2.43}$$

and the noise component ε_t is assumed to be uncorrelated with the innovation process $\boldsymbol{\nu}_s$ at all times:

$$\text{Cov}(\varepsilon_t, \boldsymbol{\nu}_s) = \mathbf{0} \text{ for all } s, t. \tag{2.44}$$

This model complexity increases significantly compared to the simple stochastic mean model, resembling the linear state-space structures assumed in Kalman Filter frameworks.

2.3.2 Causal Estimator Formulation

The derivation of the causal GLS estimator for $\boldsymbol{\beta}_t$ is extensive and thus omitted here for brevity. Interested readers may request detailed derivations from the author.

The estimator is presented as:

$$\widehat{\boldsymbol{\beta}}_t^{GLS} = \bar{\boldsymbol{\beta}} + (\widetilde{X}_{t,t-k}^\top \Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}^{-1} \widetilde{X}_{t,t-k})^{-1} (\widetilde{X}_{t,t-k}^\top \Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}^{-1} \widetilde{\boldsymbol{y}}_{t,t-k}), \tag{2.45}$$

where the adjusted variables $\widetilde{\boldsymbol{y}}_{t,t-k}$ and $\widetilde{X}_{t,t-k}$ account for the mean structure of $\boldsymbol{\beta}_t$:

$$\widetilde{\boldsymbol{y}}_{t,t-k} = \boldsymbol{y}_{t,t-k} - X_{t,t-k}\bar{\boldsymbol{\beta}}, \tag{2.46}$$

$$\widetilde{X}_{t,t-k} = \begin{bmatrix} \boldsymbol{x}_t^\top A^0 \\ \boldsymbol{x}_{t-1}^\top A^{-1} \\ \vdots \\ \boldsymbol{x}_{t-k}^\top A^{-k} \end{bmatrix}, \tag{2.47}$$

and the variance-covariance matrix $\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}$ is constructed to reflect the uncertainty associated with both the noise ε_t and the innovations $\boldsymbol{\nu}_t$:

$$\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t} = \Sigma_\varepsilon^{t,t-k} + \text{diag}(\widetilde{X}_{t,t-k}^\top) \Sigma_\nu^{Aug} \text{diag}(\widetilde{X}_{t,t-k})^\top. \tag{2.48}$$

The block matrices $\Sigma_\varepsilon^{t,t-k}$ and Σ_ν^{Aug} encapsulate the variances and covariances of the noise and innovation processes, respectively, across different time lags. These are critical in constructing $\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}$, which informs the GLS estimator, enabling the recovery of $\boldsymbol{\beta}_t$ from the

observed data $y_t,\ \boldsymbol{x}_t$ with consideration for the underlying VAR dynamics.

$$\Sigma_{\varepsilon}^{t,t-k} = \begin{bmatrix} \sigma_{\varepsilon_t}^2 & \sigma_{\varepsilon_t,\varepsilon_{t-1}} & \cdots & \sigma_{\varepsilon_t,\varepsilon_{t-k}} \\ \sigma_{\varepsilon_{t-1},\varepsilon_t} & \sigma_{\varepsilon_{t-1}}^2 & \cdots & \sigma_{\varepsilon_{t-1},\varepsilon_{t-k}} \\ \vdots & \vdots & & \vdots \\ \sigma_{\varepsilon_{t-k},\varepsilon_t} & \sigma_{\varepsilon_{t-k},\varepsilon_{t-1}} & \cdots & \sigma_{\varepsilon_{t-k}}^2 \end{bmatrix}_{(k+1)\times(k+1)} \tag{2.49}$$

$$\operatorname{diag}(\widetilde{\boldsymbol{x}}_{t,t-k}^{\top}) = \begin{bmatrix} \boldsymbol{x}_t^{\top}A^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}_{t-1}^{\top}A^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{x}_{t-k}^{\top}A^{-k} \end{bmatrix}_{(k+1)\times(kd+d)} \tag{2.50}$$

$$\Sigma_{\nu}^{Aug} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_\nu & \Sigma_\nu & \Sigma_\nu & \cdots & \Sigma_\nu \\ 0 & \Sigma_\nu & \Sigma_\nu + A\Sigma_\nu A^{\top} & \Sigma_\nu + A\Sigma_\nu A^{\top} & \cdots & \Sigma_\nu + A\Sigma_\nu A^{\top} \\ 0 & \Sigma_\nu & \Sigma_\nu + A\Sigma_\nu A^{\top} & \begin{smallmatrix}\Sigma_\nu + A\Sigma_\nu A^{\top} + \\ A^2\Sigma_\nu (A^2)^{\top}\end{smallmatrix} & \cdots & \begin{smallmatrix}\Sigma_\nu + A\Sigma_\nu A^{\top} + \\ A^2\Sigma_\nu (A^2)^{\top}\end{smallmatrix} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \Sigma_\nu & \Sigma_\nu + A\Sigma_\nu A^{\top} & \begin{smallmatrix}\Sigma_\nu + A\Sigma_\nu A^{\top} + \\ A^2\Sigma_\nu (A^2)^{\top}\end{smallmatrix} & \cdots & \begin{smallmatrix}\Sigma_\nu + A\Sigma_\nu A^{\top} + \\ \cdots + A^k\Sigma_\nu (A^k)^{\top}\end{smallmatrix} \end{bmatrix}_{(kd+d)\times(kd+d)} \tag{2.51}$$

2.3.3 Non-Causal Estimator Formulation

Again, thanks to the inherent symmetry in time series, the GLS estimator is

$$\widehat{\boldsymbol{\beta}}_t^{GLS} = \bar{\boldsymbol{\beta}} + (\widetilde{X}_{t-k,t+k}^{\top}\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}^{-1}\widetilde{X}_{t-k,t+k})^{-1}(\widetilde{X}_{t-k,t+k}^{\top}\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}^{-1}\widetilde{\boldsymbol{y}}_{t-k,t+k}), \tag{2.52}$$

where all vectors are concatenated vectors in the causal estimator $v^{nc} = \begin{bmatrix} v[k:2] & v[1] & v[2:k] \end{bmatrix}$, and matrices being a block-diagonal format with elements in the causal estimator, where $A^{nc} = \text{blkdiag}\{A[k:2, k:2],\ A[1,1],\ A[2:k, 2:k]\}$.

2.4 General Stochastic Beta Model

2.4.1 Model Set-up

The General Stochastic Beta Model presents a versatile framework for capturing the dynamic relationship between the observed variable y_t and a set of explanatory variables $\boldsymbol{x}_t$. The model structure is given by:

$$y_t = \boldsymbol{x}_t^{\top}\boldsymbol{\beta}_t + \varepsilon_t, \tag{2.53}$$

$$\text{where } \mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}|\boldsymbol{X}] = \Sigma_{\varepsilon}. \tag{2.54}$$

Here, $\boldsymbol{\beta}_t$ is defined as a stationary process with a well-specified covariance structure:

$$\boldsymbol{\beta}_t \text{ is stationary with } \mathbb{E}[(\boldsymbol{\beta}_{t+k} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}})^{\top}] = \Gamma_k. \tag{2.55}$$

By Wold's Decomposition Theorem, this represents all possible stationary processes, affirming the model's suitability for a wide range of stochastic processes.

Again, I assume that the noise ε_t is uncorrelated with all future and past innovations in $\boldsymbol{\beta}_t$:

$$\text{Cov}(\varepsilon_t, \boldsymbol{\nu}_s) = \mathbf{0} \text{ for all } s, t. \tag{2.56}$$

2.4.2 Causal Estimator Formulation

The GLS estimator for $\boldsymbol{\beta}_t$ is a testament to the model's adaptability and encompasses a broad spectrum of dynamic behaviors:

$$\widehat{\boldsymbol{\beta}}_t^{GLS} = \bar{\boldsymbol{\beta}} + (X_{t,t-k}^\top \Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}^{-1} X_{t,t-k})^{-1} (X_{t,t-k}^\top \Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}^{-1} \widetilde{\boldsymbol{y}}_{t,t-k}), \tag{2.57}$$

where the modified response vector $\widetilde{\boldsymbol{y}}_{t,t-k}$ and the design matrix $X_{t,t-k}$ are adjusted for the mean structure of $\boldsymbol{\beta}_t$:

$$\widetilde{\boldsymbol{y}}_{t,t-k} = \boldsymbol{y}_{t,t-k} - X_{t,t-k}\bar{\boldsymbol{\beta}}, \tag{2.58}$$

$$X_{t,t-k} = \begin{bmatrix} \boldsymbol{x}_t^\top \\ \boldsymbol{x}_{t-1}^\top \\ \vdots \\ \boldsymbol{x}_{t-k}^\top \end{bmatrix}. \tag{2.59}$$

The variance-covariance matrix $\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t}$ integrates the noise and the stationary properties of $\boldsymbol{\beta}_t$, ensuring that the estimator captures the full spectrum of potential correlations:

$$\Omega_{\boldsymbol{u}|\boldsymbol{\beta}_t} = \Sigma_\varepsilon^{t,t-k} + \text{diag}(X_{t,t-k}^\top) \Sigma_\nu^{Aug} \text{diag}(X_{t,t-k})^\top. \tag{2.60}$$

The matrices $\Sigma_\varepsilon^{t,t-k}$ and Σ_ν^{Aug} are meticulously constructed to reflect the covariance of the noise and the auto-covariance of $\boldsymbol{\beta}_t$, respectively. Intuitively,

$$\Sigma_\varepsilon^{t,t-k} = \text{Covariance matrix of } \boldsymbol{\varepsilon} \text{ over time}, \tag{2.61}$$

$$\Sigma_\nu^{Aug} = \text{Augmented covariance matrix capturing the dynamics of } \boldsymbol{\beta}_t. \tag{2.62}$$

More tediously, the exact mathematical formulae are

$$\Sigma_{\varepsilon}^{t,t-k} = \begin{bmatrix} \sigma_{\varepsilon_t}^2 & \sigma_{\varepsilon_t,\varepsilon_{t-1}} & \cdots & \sigma_{\varepsilon_t,\varepsilon_{t-k}} \\ \sigma_{\varepsilon_{t-1},\varepsilon_t} & \sigma_{\varepsilon_{t-1}}^2 & \cdots & \sigma_{\varepsilon_{t-1},\varepsilon_{t-k}} \\ \vdots & \vdots & & \vdots \\ \sigma_{\varepsilon_{t-k},\varepsilon_t} & \sigma_{\varepsilon_{t-k},\varepsilon_{t-1}} & \cdots & \sigma_{\varepsilon_{t-k}}^2 \end{bmatrix}_{(k+1)\times(k+1)} \tag{2.63}$$

$$\operatorname{diag}(\boldsymbol{x}_{t,t-k}^{\top}) = \begin{bmatrix} \boldsymbol{x}_t^{\top} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}_{t-1}^{\top} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{x}_{t-k}^{\top} \end{bmatrix}_{(k+1)\times(kd+d)} \tag{2.64}$$

$$\Sigma_{\nu}^{Aug} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 2\Gamma_0 - \Gamma_1^{\top} - \Gamma_1 & \begin{smallmatrix}\Gamma_0-\Gamma_1^{\top}\\ +\Gamma_1-\Gamma_2\end{smallmatrix} & \cdots & \begin{smallmatrix}\Gamma_0-\Gamma_1^{\top}\\ +\Gamma_{k-1}-\Gamma_k\end{smallmatrix} \\ 0 & \begin{smallmatrix}\Gamma_0-\Gamma_1\\ +\Gamma_1^{\top}-\Gamma_2^{\top}\end{smallmatrix} & 2\Gamma_0 - \Gamma_2^{\top} - \Gamma_2 & \cdots & \begin{smallmatrix}\Gamma_0-\Gamma_2^{\top}\\ +\Gamma_{k-2}-\Gamma_k\end{smallmatrix} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \begin{smallmatrix}\Gamma_0-\Gamma_1\\ +\Gamma_{k-1}^{\top}-\Gamma_k^{\top}\end{smallmatrix} & \begin{smallmatrix}\Gamma_0-\Gamma_2\\ +\Gamma_{k-2}^{\top}-\Gamma_k^{\top}\end{smallmatrix} & \cdots & 2\Gamma_0 - \Gamma_k^{\top} - \Gamma_k \end{bmatrix}_{(kd+d)\times(kd+d)} \tag{2.65}$$

This GLS estimator not only highlights the model's broad applicability across different temporal structures, but also its precision in estimating the intricate intertemporal relationships inherent in the stochastic processes governing $\boldsymbol{\beta}_t$.

Chapter 3

Empirical Methodology

Chapter 3 presents a detailed exploration of the feasible application of the theoretical models developed in the previous chapter. It begins with Section 3.1, where the primary aims and objectives of the empirical methodologies employed in this study are clearly outlined.

This is followed by Section 3.2, which is subdivided into two parts: Section 3.2.1, 'Feasible GLS,' delves into the specifics of the Feasible Generalized Least Squares method, its implementation, and its relevance to the study. Section 3.2.2, 'Optimal Window Length for OLS/WLS,' discusses the determination of the optimal window length for Ordinary Least Squares and Weighted Least Squares, a critical factor in enhancing the accuracy of parameter estimation.

The chapter concludes with Section 3.3, where the development and application of an iterative estimation method are discussed, demonstrating its effectiveness in refining the estimation process.

3.1 Objective

Parameter Uncertainty

In this empirical analysis, I confront the challenge of parameter uncertainty, focusing on the estimation of σ_ν^2 and σ_ε^2 as a motivating example. These parameters are central to our model but their true values remain unknown, leading to complexities in our estimation approach. This uncertainty is not just a mathematical inconvenience; it has substantive implications for the way we interpret and apply our model results.

Consequences of Parameter Uncertainty

The presence of parameter uncertainty necessitates the adoption of a Feasible Generalized Least Squares (fGLS) approach instead of the standard GLS method. Feasible GLS takes into account the estimated nature of these parameters, adjusting the estimation technique to mitigate potential biases and inaccuracies.

Additionally, parameter uncertainty complicates the determination of an optimal estimation window size, particularly for rolling OLS and WLS. Choosing the right window size is a delicate balance – too small, and the model may miss out on significant trends; too large,

and it may become insensitive to recent changes. This selection process becomes increasingly challenging without a clear understanding of the true parameter values, necessitating a careful, data-driven approach to window size optimization.

Overall, the objective of this empirical methodology is to navigate these uncertainties and complexities to provide robust, reliable estimations that can withstand scrutiny and contribute meaningful insights to our field of study.

3.2 Method-of-Moment Estimator

The Method-of-Moment (MoM) Estimator is a fundamental statistical technique used to estimate parameters in models where direct measurement isn't feasible. It calculates parameter values by equating sample moments (like means, and variances) with their theoretical counterparts. MoM is particularly relevant in this context for estimating parameters like σ_ν^2 and σ_ε^2, which are essential but unobservable in my model. By employing MoM, these parameters can be indirectly estimated from the observable data, thus enabling the application of feasible GLS methods in my analysis. The utilization of MoM in this scenario facilitates a more accurate and practical approach to handling parameter uncertainty in my empirical methodology.

3.2.1 Feasible GLS

Estimation of $\boldsymbol{\sigma_\nu^2}$ and $\boldsymbol{\sigma_\varepsilon^2}$

Considering our earlier equations:

$$y_t = \mu_t + \varepsilon_t, \tag{3.1}$$

$$y_{t-1} = \mu_t - \nu_t + \varepsilon_{t-1}, \tag{3.2}$$

Let's take the first differences for both equations:

$$\Delta y_t = \nu_t + \varepsilon_t - \varepsilon_{t-1}, \tag{3.3}$$

$$\Delta y_{t-1} = \nu_{t-1} + \varepsilon_{t-1} - \varepsilon_{t-2}. \tag{3.4}$$

Analyzing the autocovariance function (ACVF) of (Δy_t), we obtain:

$$\gamma_0 = \sigma_\nu^2 + 2\sigma_\varepsilon^2, \tag{3.5}$$

$$\gamma_1 = -\sigma_\varepsilon^2. \tag{3.6}$$

This leads to the estimates:

$$\widehat{\sigma}_\varepsilon^2 = -\widehat{\gamma}_1(\Delta y_t), \tag{3.7}$$

$$\widehat{\sigma}_\nu^2 = \widehat{\gamma}_0(\Delta y_t) + 2\widehat{\gamma}_1(\Delta y_t). \tag{3.8}$$

The feasible GLS estimator is then given by:

$$\widehat{\mu}_t^{fGLS} = (\iota^\top \widehat{\Omega}^{-1}_{\boldsymbol{u}|\mu_t} \iota)^{-1} (\iota^\top \widehat{\Omega}^{-1}_{\boldsymbol{u}|\mu_t} \boldsymbol{y}_{t,t-k}). \tag{3.9}$$

This section highlights the practical application of the MoM Estimator in the context of feasible GLS, elucidating the process of estimating critical parameters and subsequently applying them in the GLS estimator.

3.2.2 Optimal Window Length for OLS/WLS

The determination of the optimal window size, denoted as k^*, is crucial for ensuring the accuracy of the naive rolling OLS or WLS estimators.

Under the correct model specification, the difference in consecutive estimates of μ_t should converge in distribution to the true underlying distribution, with variance:

$$\text{Var}[\Delta\widehat{\mu}_t] \xrightarrow{p} \sigma_\nu^2, \tag{3.10}$$

, where σ_ν^2 represents the variance of the state innovations. The optimal window horizon k^* can be chosen such that it minimizes the difference between the estimated variance of $\Delta\widehat{\mu}_t$ and the estimated variance parameter $\widehat{\sigma}_\nu^2$. Formally, this can be expressed as:

$$k^* = \text{argmin}\left\{\left|\widehat{\text{Var}}[\Delta\widehat{\mu}_t] - \widehat{\sigma}_\nu^2\right|\right\}, \tag{3.11}$$

, where $\widehat{\sigma}_\nu^2$ is estimated by the method of moments as previously discussed.

This approach ensures that the chosen window length aligns closely with the underlying statistical properties of the time series data, thereby enhancing the accuracy and reliability of my estimations.

3.3 Iterative Estimator

Rationale for an Iterative Approach

The ratio $\sigma_\varepsilon^2/\sigma_\nu^2$ is a significant factor in the model's accuracy. In scenarios where σ_ν^2 is small, an overestimation or underestimation of $\widehat{\sigma}_\nu^2$ can lead to substantial discrepancies in the model outcomes. Therefore, a more refined approach to estimate σ_ν^2 is imperative.

Alternative Estimation and Iterative Refinement

This methodology is motivated by the two-step estimation of the raw GLS model. In this section, I would like to propose an alternative estimation method for $\sigma_\varepsilon^2/\sigma_\nu^2$, where the ratio is estimated by the calculated variance from the first-step estimate of $\widehat{\mu}_t$:

$$\frac{\widetilde{\sigma}_\varepsilon^2}{\widetilde{\sigma}_\nu^2} = \frac{\widehat{\text{Var}}[y_t - \widehat{\mu}_t]}{\widehat{\text{Var}}[\Delta\widehat{\mu}_t]} \tag{3.12}$$

When the window size k is large enough, it yields a more stable estimation of $\sigma_\varepsilon^2/\sigma_\nu^2$. This stability is crucial, especially in models where parameter changes are subtle yet significant. Empirically, I show that this method produces estimates of$\widetilde{\sigma}_\varepsilon^2/\widetilde{\sigma}_\nu^2$ that are consistently less biased compared to the initial model input of $\widehat{\sigma}_\varepsilon^2/\widehat{\sigma}_\nu^2$.

Iterative Procedure and Model Adaptability

The iterative procedure involves substituting this newly estimated $\widetilde{\sigma}_\varepsilon^2/\widetilde{\sigma}_\nu^2$ back into the feasible GLS estimator, re-estimating $\widehat{\mu}_t$, and re-evaluating $\widetilde{\sigma}_\varepsilon^2/\widetilde{\sigma}_\nu^2$. This process is repeated,

further refining the estimate of the variance ratio. The adaptability of this iterative method makes it a robust tool, particularly for models where traditional estimators may not perform optimally. It exemplifies an approach where the model dynamically adjusts to new information, enhancing both the precision and reliability of the estimations.

In addition, this methodology can be easily adapted to fit more complex multi-variate settings in the estimation of the state model or measurement equation. Its performance will be examined in Section 4.3.

Chapter 4

Simulation

Chapter 4 undertakes a detailed comparative analysis of various models under different scenarios, providing insights into their performance and applicability. The chapter commences with Section 4.1, where the framework for the simulations is established. This section details the scenarios used for comparison, including known and unknown parameters and the differentiation between causal and non-causal filters.

Following this, Section 4.2 evaluates the performance of several models, such as Rolling OLS, Rolling WLS, the Kalman Filter, and the proposed Rolling GLS Model, with predetermined parameter values. This section is instrumental in demonstrating the effectiveness of these models under controlled conditions, offering a clear benchmark for their performance.

4.1 Simulation Set-up

The simulation aims to compare various models under four distinct scenarios: {Known parameters, Unknown parameters} $\times$ {Causal filters, Non-causal filters}.

Causal Filter Comparison

In the realm of causal filters, the following models were compared:

1. Rolling Ordinary Least Squares (OLS),
2. Rolling Linear Decaying Weighted Least Squares (WLS),
3. Kalman Filter,
4. Proposed Rolling Generalized Least Squares (GLS) Model.

Non-causal Filter Comparison

For the non-causal setting, the models included:

1. Rolling OLS,
2. Rolling Linear Decaying WLS,

3. Kalman Smoother,

4. Hodrick-Prescott Filter,

5. Baxter-King Filter,

6. Proposed Rolling GLS Model.

The chosen model for simulation is a homoskedastic stochastic mean model, described by:

$$y_t = \mu_t + \varepsilon_t, \tag{4.1}$$
$$\mu_t = \mu_{t-1} + \nu_t, \tag{4.2}$$
$$\varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2), \tag{4.3}$$
$$\nu_t \sim \mathcal{WN}(0, \sigma_\nu^2), \tag{4.4}$$
$$\mathrm{Cov}(\varepsilon_t, \nu_s) = 0 \text{ for all } s, t. \tag{4.5}$$

This simple model was chosen for its compatibility with earlier naive models, like the Hodrick-Prescott Filter.

4.2 Known Parameters

4.2.1 Causal Filters

In the simulation under the scenario of known parameters, I evaluated the performance of various models with predetermined parameter values. These parameters were set as $\sigma_\nu^2 = 0.1$, $\sigma_\varepsilon^2 = 1$, and a total of $T = 50,000$ observations.

Traditionally, in the realm of time-varying parameter estimation, the selection of an appropriate window size is critical and is subject to a bias-variance trade-off. A window that is too short may lead to an estimator that is unduly influenced by noise, as it captures only the volatility of a small subset of data, mistaking random fluctuations for genuine variations. Conversely, an overly extended window size can introduce significant bias, as it incorporates stale data points that no longer reflect the current dynamics of the parameter, leading to estimations that are systematically skewed.

However, due to the optimized weighting scheme for my proposed GLS-based model, this is no longer a problem. One is free to choose an arbitrarily long estimation window, and my model will still give the optimal estimations.

The results are shown in Figure 4.1 and Table 4.1. To begin with, for the rolling OLS (EW) and rolling WLS (EW) model, we observe from the figure that the RMSE tends to decrease and then increase, indicating a U-shaped relationship between window size and estimation accuracy. This is consistent with our theory about the bias-variance trade-off.

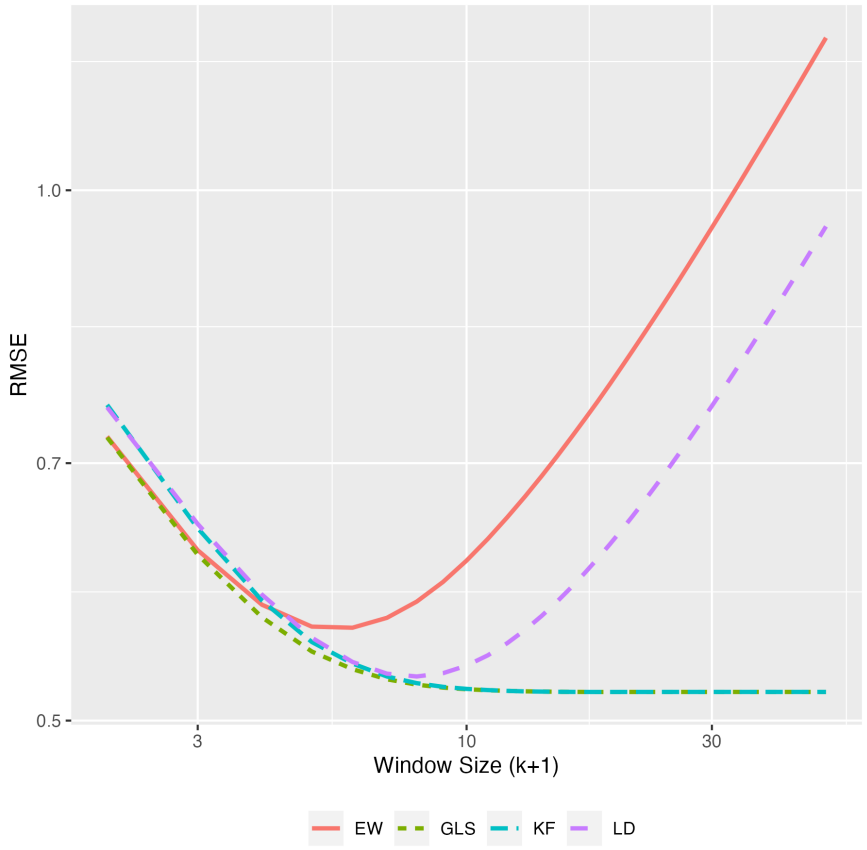


Figure 4.1: RMSE Comparison Across Causal Filters with Known Parameters. This figure illustrates the relationship between the window size and RMSE for various causal filters. The curves represent different filters: 'EW' for Rolling OLS, 'LD' for Rolling WLS, 'KF' for the Kalman Filter, and 'GLS' for the Proposed Rolling GLS Model. The x-axis denotes the window size $(k+1)$, and the y-axis represents the RMSE.

Table 4.1: Performance Comparison of Causal Filters. The table summarizes the optimal window length and the corresponding RMSE percentage for each filter. Notations used are 'EW' for Rolling OLS, 'LD' for Rolling WLS, 'KF' for the Kalman Filter, and 'GLS' for the Proposed Rolling Generalized Least Squares model. An infinite window length (∞) indicates that the model utilizes all available data points.

Model	Alias	Optimal Window Length	Optimal RMSE (%)
Rolling OLS	EW	6	57.06
Rolling WLS	LD	8	53.45
Kalman Filter	KF	∞	52.37
Proposed Model	GLS	∞	52.37

In addition, Figure 4.1 also indicates that the proposed GLS-based filter outperforms other models across varying window sizes. It maintains a consistently low RMSE, demon-

strating robustness and efficiency in parameter estimation. It is true that the Kalman Filter (KF) exhibits performance comparable to the GLS approach, but it still underperforms with smaller window sizes. I suspect this underperformance is related with the lack of optimization when initializing the priors $\mu_{1|0}$ and the variance $P_{1|0}$.

This robustness to varying window sizes is particularly advantageous in macroeconomic studies, where data availability and market conditions can fluctuate significantly. By offering optimal estimations regardless of the window length, the GLS model circumvents the traditional bias-variance dilemma, providing a versatile and reliable tool for analysts dealing with diverse and dynamic datasets.

In summary, the stability and lower RMSE of the GLS method across all window sizes suggest it is the superior model in this simulation.

4.2.2 Non-Causal Filters

My simulation then explores the efficacy of various non-causal filters. The lambda and frequency band settings for the Hodrick-Prescott Filter and Baxter-King Filter are cross-validated so that they have the best out-of-sample (OOS) performance. The provided figure, Figure 4.2, shows the Root Mean Square Error trends across different models as the window size increases. Table 4.2 summarises the performance metrics.

With the parameters set to $\sigma_\nu^2 = 0.1$ and $\sigma_\varepsilon^2 = 1$, we see a 13% decrease in value (or 24% decrease in relative percentage) of RMSE by moving to the non-casual two-sided filters. The proposed GLS model and the Kalman Smoother (KS) still outperform other models, achieving the lowest RMSE, which underscores the effectiveness of these methods in handling parameter uncertainty. Again, Kalman Smoother demonstrates inferior performance compared to the proposed model for small windows, thanks to the sub-optimal initialization.

The Hodrick-Prescott (HP) and Baxter-King (BK) filters show competitive performance, yet they do not reach the efficiency of the GLS and KS methods, even for the selected "optimal" lambda and frequency band settings after cross-validation.

Table 4.2: Performance of Non-Causal Filters. This table presents the optimal window length and RMSE for different non-causal filters. The aliases are 'EW' for Rolling OLS, 'LD' for Rolling WLS, 'BK' for Baxter-King Filter, 'HP' for Hodrick-Prescott Filter, 'KS' for Kalman Smoother, and 'GLS' for the Proposed Model. The notation ∞ signifies an infinite window length, implying the use of all data points.

Model	Alias	Optimal Window Length	Optimal RMSE (%)
Rolling OLS	EW	11	43.01
Rolling WLS	LD	15	40.64
Baxter-King Filter	BK	∞	41.40
Hodrick-Prescott Filter	HP	∞	41.01
Kalman Smoother	KS	∞	39.91
Proposed Model	GLS	∞	39.91

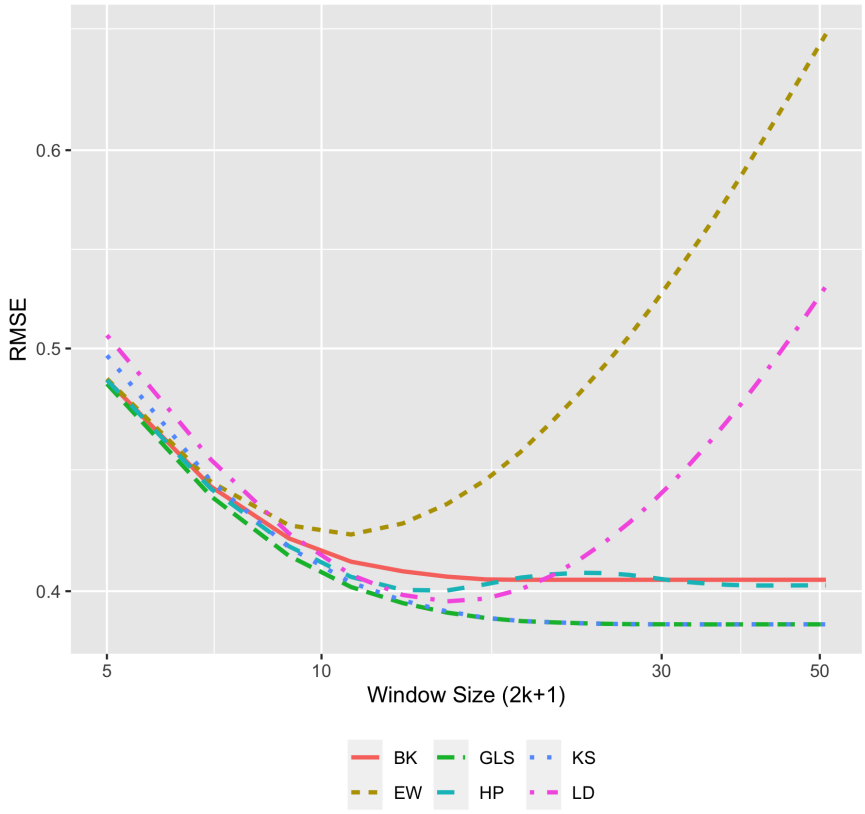


Figure 4.2: RMSE Across Non-Causal Filters with Known Variances. The figure depicts the RMSE as a function of window size for various non-causal filters. Alias notations are 'EW' for Rolling OLS, 'LD' for Rolling WLS, 'BK' for Baxter-King Filter, 'HP' for Hodrick-Prescott Filter, 'KS' for Kalman Smoother, and 'GLS' for the Proposed Model. The window size is given by $k + 1$, and the RMSE is measured on the y-axis.

4.2.3 Robustness to Log-Normal Noise

In this part of the simulation, I assess the robustness of the proposed GLS model compared to the Kalman Filter when the underlying noise distribution deviates from normality. Specifically, I introduce log-normal noise, re-centered to ensure a mean of zero, to investigate the models' performances under a more realistic and challenging data distribution.

The GLS model, grounded in the Generalized Method of Moments, boasts flexibility by relying solely on some moment conditions and is not constrained by distributional assumptions. This contrasts with the Kalman Filter, which typically presupposes joint normality for the measurement and state innovation errors—a condition that may not hold in empirical data.

Figure 4.3 illustrates the comparative performance of these models with the introduction of log-normal noise. The GLS model demonstrates enhanced robustness, outperforming the Kalman Filter, particularly in small samples. This suggests a marked superiority of the GLS model, evidenced by a lower Root Mean Square Error (RMSE) across different window sizes.

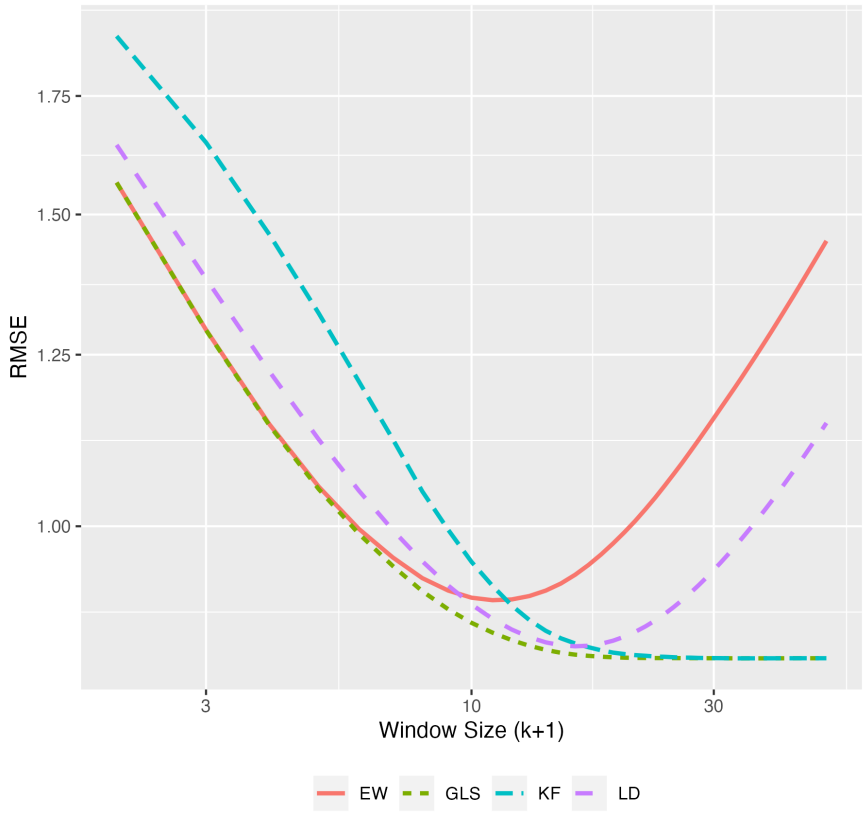


Figure 4.3: RMSE Comparison of Causal Filters under Log-Normal Noise. This figure demonstrates the RMSE performance of various causal filters when subjected to log-normal noise, highlighting the robustness of the models under more realistic data distribution scenarios. The filters are denoted as 'EW' for Rolling OLS, 'GLS' for the Proposed GLS Model, 'KF' for the Kalman Filter, and 'LD' for Rolling WLS. The x-axis specifies the window size $(k+1)$, and the y-axis measures the RMSE.

4.3 Unknown Parameters

4.3.1 MoM Estimator-Scenario One

In our simulation exploring the scenario with unknown parameters, various models were assessed based on their ability to estimate parameters that were not directly observable. With $\sigma_\nu^2 = 0.1$, $\sigma_\varepsilon^2 = 1$, and $T = 50,000$, the models were tested for their estimation precision.

Using the autocovariance function of Δy_t, the estimated values were $\widehat{\sigma}_\nu^2 = 0.11$ and $\widehat{\sigma}_\varepsilon^2 = 0.99$. To discern the optimal window size for OLS and WLS estimators, I examined the variance of the changes in the estimated μ_t across a spectrum of window sizes k. The results, depicted in Figure 4.4, suggested optimal window lengths for each model, as captured in the subsequent table. Notably, the infinite window size for the fGLS and GLS models implies a reliance on all available data, underpinning the models' robustness to variations in window length.

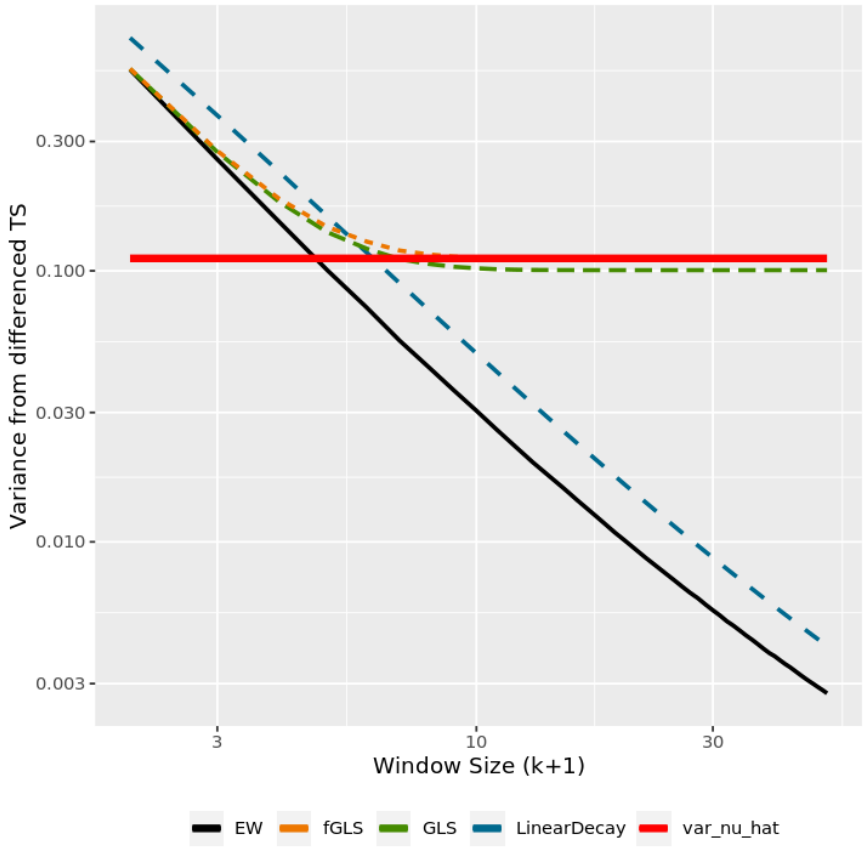


Figure 4.4: Variance of $\Delta\widehat{\mu}_t$ Across Different Window Sizes. This figure plots the variance of the estimated changes in μ_t against window sizes, highlighting the models' sensitivity to the window size in estimation. The legends 'EW', 'fGLS', 'GLS', 'LinearDecay', and 'var_nu_hat' correspond to Rolling OLS, Proposed Empirical Model, Proposed Theoretical Model, a linear decaying approach, and the variance of re-estimated noise, respectively.

The estimated optimal window sizes for rolling OLS and rolling WLS models are summarized in Table 4.3 below.

Table 4.3: Estimated Optimal Window Size with Unknown Parameters. This table reports the estimated optimal window size (k+1) for each model along with the associated RMSE. 'EW' represents Rolling OLS, 'LD' stands for Rolling WLS, 'fGLS' is the Proposed Empirical Model, and 'GLS' represents the Proposed Theoretical Model. The infinite symbol (∞) indicates the use of all data points for estimation.

Model	Alias	Estimated Optimal Window Size (k+1)
Rolling OLS	EW	5
Rolling WLS	LD	7
Proposed Empirical Model	fGLS	∞
Proposed Theoretical Model	GLS	∞

Subsequently, a comprehensive comparison of model performance was conducted. The RMSE across models was illustrated in Figure 4.5, which presents a visual summary of the predictive accuracy.

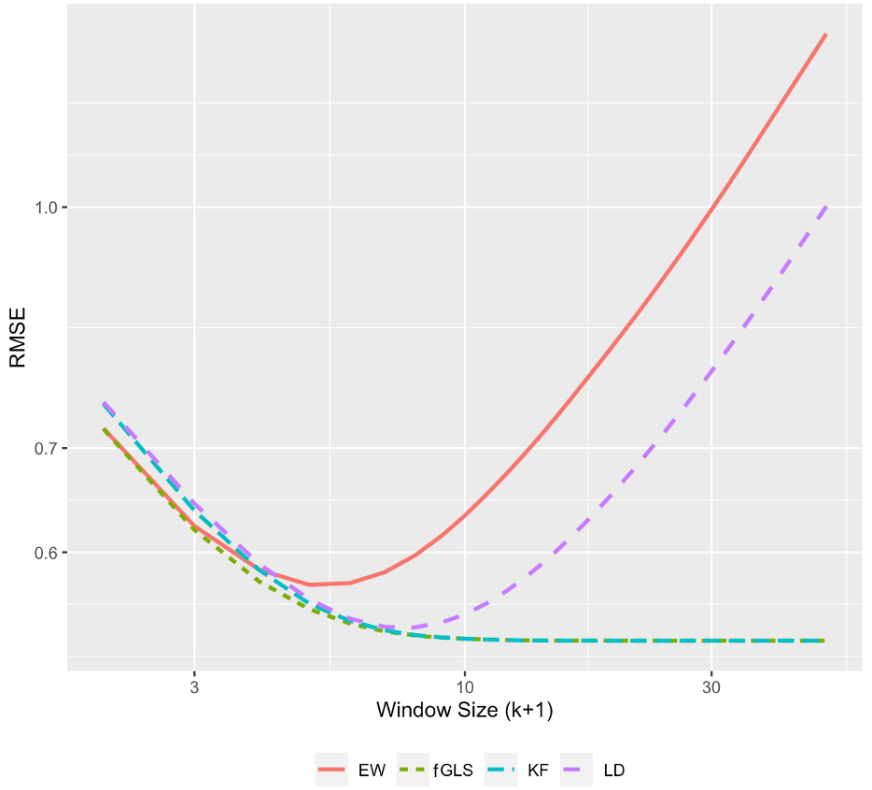


Figure 4.5: RMSE Comparison Across Models with Unknown Parameters. This figure compares the RMSE for different estimation models under the scenario of unknown parameters. The legends 'EW', 'fGLS', 'KF', and 'LD' denote Rolling OLS, Proposed Empirical Model, Kalman Filter, and Rolling WLS, respectively. The x-axis shows the window size $(k+1)$, and the y-axis shows the RMSE.

Table 4.4: Performance of Models with Unknown Parameters. The table presents the estimated optimal window size and the RMSE for models when parameters are unknown. Aliases used include 'EW' for Rolling OLS, 'LD' for Rolling WLS, 'KF' for the Kalman Filter, 'fGLS' for the Proposed Empirical Model, and 'GLS' for the Proposed Theoretical Model. The symbol ∞ denotes an infinite window size.

Model	Alias	Estimated Optimal Window Size (k+1)	RMSE (%)
Rolling OLS	EW	5	57.18
Rolling WLS	LD	7	53.64
Kalman Filter	KF	∞	52.43
Proposed Empirical Model	fGLS	∞	52.43
Proposed Theoretical Model	GLS	∞	52.37

The proposed feasible GLS model still stochastically dominates all other models examined, for all window sizes. What's more, the RMSE of the feasible model (52.43%) is very close to its theoretical optimum of 53.37%, a testament to its robustness in the face of parameter uncertainty.

4.3.2 MoM Estimator-Scenario Two

This scenario tested the adaptability of the models by reducing the stochastic term variance to $\sigma_\nu^2 = 0.01$ and maintaining $\sigma_\varepsilon^2 = 1$, across a broad set of $T = 50,000$ observations. The focus was on detecting the models' robustness to smaller parameter fluctuations, which are common in practical economics and financial time series analysis.

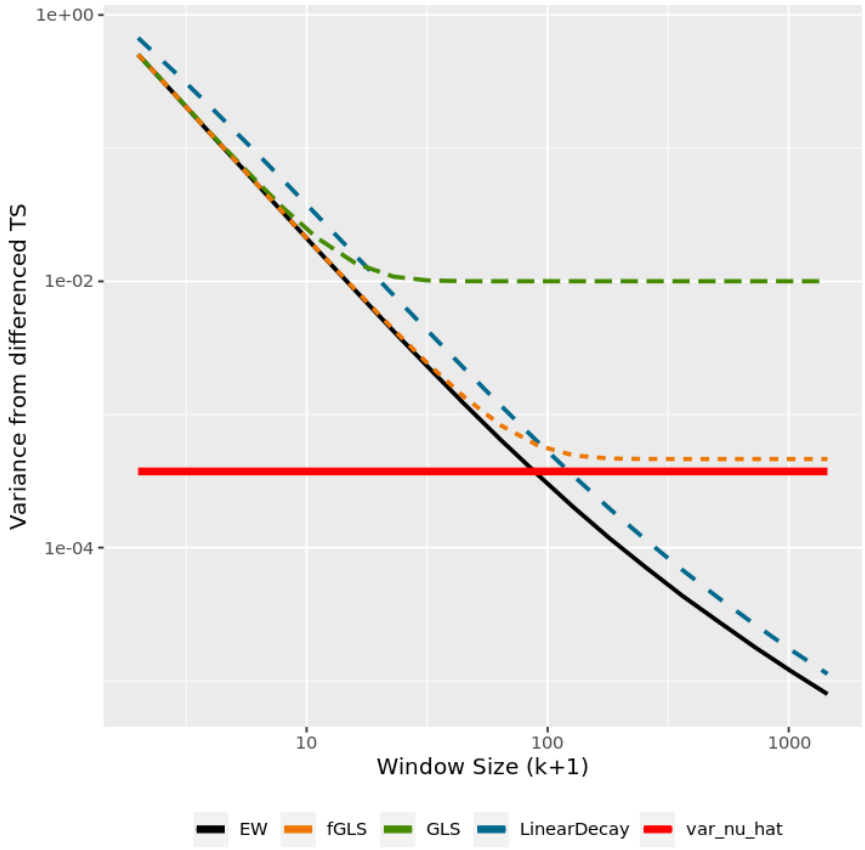


Figure 4.6: Variance of $\Delta\widehat{\mu}_t$ across Models for Scenario Two. This figure displays the variance of the estimated changes in $\widehat{\mu}_t$ as a function of window size for various models under a scenario with reduced stochastic term variance. The lines 'EW', 'fGLS', 'GLS', 'LinearDecay', and 'var_nu_hat' represent Rolling OLS, Proposed Empirical Model, Proposed Theoretical Model, a linear decay strategy, and the variance of noise re-estimation, respectively.

Figures 4.6 and 4.7 reveal the outcomes of the simulation. In Figure 4.6 and Table 4.5, the optimal window size for the OLS and WLS methodologies was evaluated, showing a considerable deviation from the theoretical benchmark. Figure 4.7 further highlights the

discrepancy, as the RMSE of the empirical model (fGLS) significantly diverges from the theoretical counterpart (GLS), more noticeably with increased window sizes.

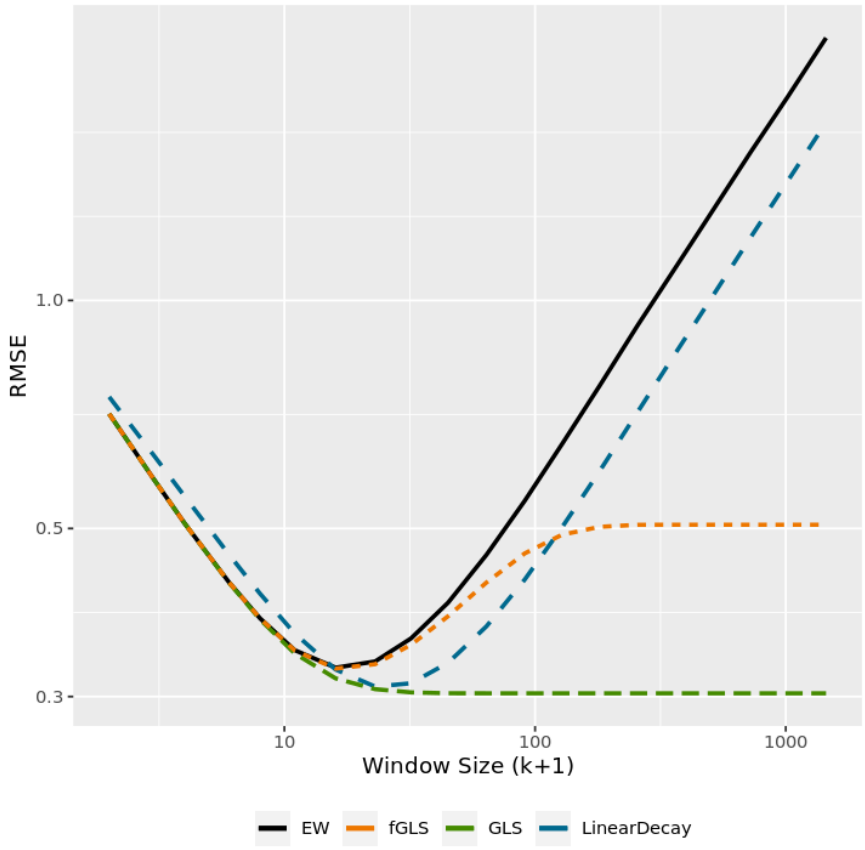


Figure 4.7: RMSE across Models for Scenario Two. The figure compares the RMSE across different estimation models when faced with a scenario of decreased variance in the stochastic term. Notations 'EW', 'fGLS', 'GLS', and 'LinearDecay' correspond to Rolling OLS, Proposed Empirical Model, Proposed Theoretical Model, and a linear decaying method, respectively. The window size is plotted on a logarithmic scale $(k+1)$ to better visualize the RMSE over a wide range of values.

Table 4.5: Performance of Empirical Model vs. Theoretical Model in Scenario Two. This table contrasts the estimated optimal window size and the best RMSE percentage obtained from the empirical model against the theoretical model. 'EW' stands for Rolling OLS, 'LD' for Rolling WLS, 'fGLS' for the Proposed Empirical Model, and 'GLS' for the Proposed Theoretical Model. Theoretical window sizes and RMSEs reflect the model's optimal performance based on known parameter values.

Alias	Est. Window	Theo. Window	Est. Best RMSE (%)	Theo. Optimal RMSE (%)
EW	88	18	53.53	32.70
LD	122	26	49.01	30.84
fGLS	∞	18	50.57	32.56
GLS	∞	∞	30.30	30.30

I observed a significant disparity between estimated and theoretical outcomes, prompting a critical evaluation of the empirical methodology's effectiveness in such a modified analytical scenario.

The primary factor contributing to the performance gap was the estimation of $\widehat{\sigma}_\nu^2$ and $\widehat{\sigma}_\varepsilon^2$. The MoM estimator calculated σ_ν^2 at an estimated 0.00037, a minute absolute error but a substantial one in relative terms. This is especially pertinent given the GLS-based filter's reliance on the ratio $\sigma_\varepsilon^2/\sigma_\nu^2$, where the actual ratio of 100 contrasts sharply with the estimated 2730. Such a disparity is at the core of the empirical models' suboptimal performance in this scenario.

To address these estimation inaccuracies, I employed an iterative estimator as outlined in Section 3.3. This approach incrementally refines the estimates of σ_ν^2 and σ_ε^2, aiming to reconcile them with their true values, thus enhancing the empirical model's precision and reliability.

4.3.3 Iterative Estimator

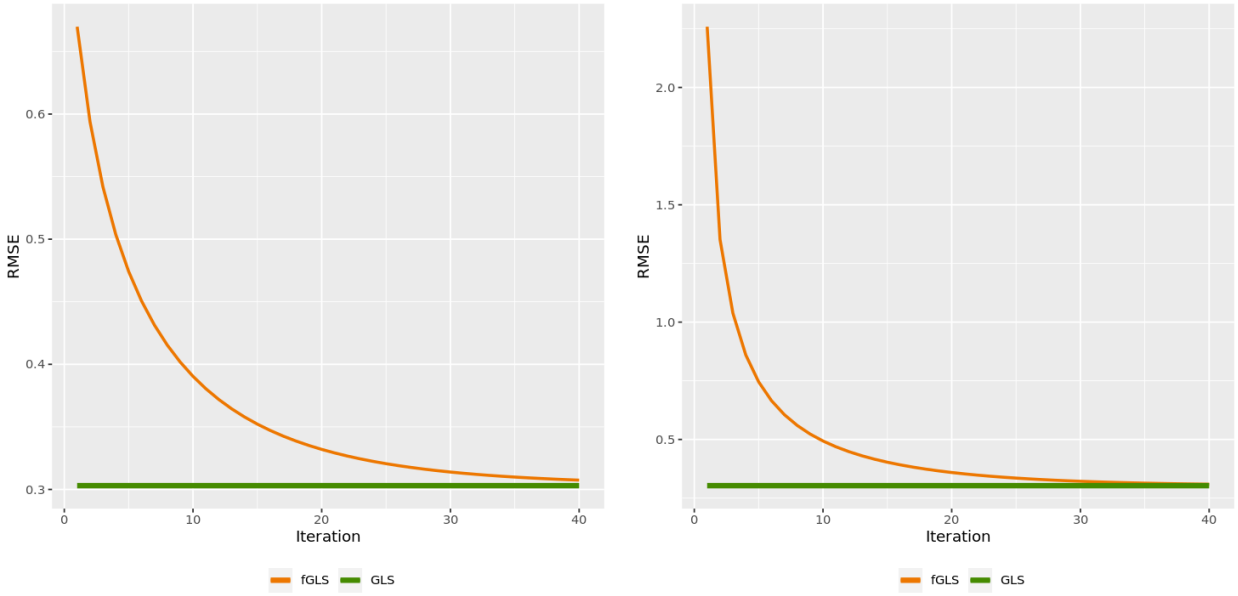
I introduce the iterative estimator to improve the precision of the critical parameters σ_ν^2 and σ_ε^2. The accuracy of these estimates is pivotal for the efficacy of the GLS-based filter. My initial estimates indicated a significant relative error, which necessitated an iterative correction mechanism to realign the estimates with their actual values.

For this simulation, set parameters were $\sigma_\nu^2 = 0.01$, $\sigma_\varepsilon^2 = 1$, spanning $T = 50,000$ observations. A substantial estimation window of $k + 1 = 1500$ was chosen to ensure a comprehensive capture of the right tail anomalies. The robustness of the iterative model was tested using markedly incorrect initial values for $\widehat{\sigma}_\nu^2$, specifically 0 and 1, in stark contrast to the true value.

Figures 4.8a and 4.8b depict the model's exponential convergence to the theoretical standard, validating the iterative estimator's effectiveness. Notwithstanding the initial values' significant deviation from the true parameters, the empirical model achieved rapid convergence, stabilizing at robust estimates within about 40 iterations.

For practical applications, I recommend initializing the parameters in the measurement and state models with the Method of Moments estimator or established Generalized Method

of Moments estimators. Then, employing the iterative techniques expounded in Section 3.3 ensures a solid foundation of the iterative process, leading to efficient and accurate parameter estimates.



(a) Iterative Estimation with $\widehat{\sigma}_{\nu}^2$ Initialized at 1 (b) Iterative Estimation with $\widehat{\sigma}_{\nu}^2$ Initialized at 0

Figure 4.8: Convergence of RMSE in the Iterative Estimator. Subfigure (a) shows the convergence of RMSE for the fGLS and GLS models with the state variance $\hat{\sigma}_v^2$ initialized at 1, and Subfigure (b) displays the convergence with $\hat{\sigma}_v^2$ initialized at 0. These graphs demonstrate the impact of initial parameter values on the convergence rate and stability of the iterative estimation process.

Chapter 5

Real-World Applications

The empirical evidence and simulations have laid the groundwork for practical applications of the proposed model. Let's now turn our attention to real-world applications, particularly in the asset pricing field.

This chapter is divided into three sections, each focusing on a different aspect of real-world application. The first section delves into the use of the models in portfolio construction, exploring their implications for optimizing portfolio strategies. The second section focuses on factor estimation, assessing the application of the models in evaluating market trends and risk factors. The final section addresses the models' utility in understanding and forecasting the inflation risk premium, a critical element in financial risk management.

Collectively, this chapter provides a comprehensive overview of the real-world relevance and applicability of the developed models, bridging the gap between theoretical statistics research and practical financial analysis.

5.1 Portfolio Construction

The endeavor begins with the task of constructing an optimal complete portfolio utilizing the classical Markowitz 1952 model. The model considers an investor's choice among a riskless asset and N risky assets. The investor seeks to maximize a mean-variance objective function, which, given the estimates of mean returns μ and covariance matrix Σ, yields the optimal portfolio weights w^*.

$$w^* = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{5.1}$$

Note that $\phi^2 = \mu'\Sigma^{-1}\mu$ is the squared Sharpe ratio of the ex-ante tangency portfolio of the risky assets.

However, in a real-world scenario, μ and Σ are not directly observable and must be estimated from historical data. This estimation is traditionally performed using a rolling window of returns data. For this study, a 60-month estimation window was employed. The variance-covariance matrix is directly estimated from data in a native plug-in approach, and the mean returns were estimated using a variety of filters including Rolling OLS, Kalman Filter, and the model proposed in this thesis.

The empirical evaluation period spanned from July 1963 to June 2023, with the initial 5 years dedicated to the estimation phase and the subsequent period serving as a backtest. The analysis was conducted on various portfolios sourced from Kenneth French's data library:

- Market Portfolio
- Three Portfolios Formed on Size
- Five Portfolios Formed on Operating Profitability
- Six Portfolios Formed on Size and Book-to-Market

The performance of the portfolios constructed using different estimation strategies was assessed based on their annualized Sharpe Ratios.

Table 5.1: Annualized Sharpe Ratios for Portfolios Using Various Estimation Strategies. This table compares the performance of portfolios constructed with different estimation methods in terms of their annualized Sharpe Ratios. The models evaluated include Rolling OLS, Kalman Filter, and the Proposed Model, across various portfolio compositions derived from Kenneth French's data library.

Composite	**Rolling OLS**	**Kalman Filter**	**Proposed Model**
Market Portfolio	0.43	0.47	0.46
Size only	0.52	0.63	0.63
Operating Profitability only	0.58	0.67	0.64
Size and Book-to-Market	0.58	0.70	0.72

The results, summarized in Table 5.1, indicate that the proposed model demonstrates a comparative advantage over the Rolling OLS approach and exhibits similar efficacy to the Kalman Filter. This suggests that the proposed model can improve portfolio performance metrics such as the Sharpe Ratio in a real-world setting.

5.2 Factor Estimation

In contemporary investment practices, accurately gauging systematic risk is of paramount importance. Section 1.2.1 discusses how a firm's systematic risk profile can shift in response to operational changes, such as adjustments in leverage that manifest as fluctuations in market beta. Traditional econometric models, along with more sophisticated machine learning approaches, often fall short of capturing these dynamic variations. This deficiency paves the way for the introduction of the model developed in this thesis, which adeptly addresses these changes.

This research applies three model configurations to estimate systematic risk:

Stochastic Mean Model:

$$r_t = \mu_t + \varepsilon_t \tag{5.2}$$

$$\mu_t = c + a\mu_{t-1} + \nu_t \tag{5.3}$$

Market Model:

$$r_t = \alpha_t + \beta_t r_t^m + \varepsilon_t \tag{5.4}$$

$$\begin{bmatrix} \beta_t \\ \alpha_t \end{bmatrix} = c + A \begin{bmatrix} \beta_{t-1} \\ \alpha_{t-1} \end{bmatrix} + \nu_t \tag{5.5}$$

Fama-French Three-Factor Model:

$$r_t = \begin{bmatrix} r_t^m & \text{SMB}_t & \text{HML}_t & 1 \end{bmatrix} \begin{bmatrix} \beta_t \\ s_t \\ h_t \\ \alpha_t \end{bmatrix} + \varepsilon_t \tag{5.6}$$

$$\begin{bmatrix} \beta_t \\ s_t \\ h_t \\ \alpha_t \end{bmatrix} = c + A \begin{bmatrix} \beta_{t-1} \\ s_{t-1} \\ h_{t-1} \\ \alpha_{t-1} \end{bmatrix} + \nu_t \tag{5.7}$$

Each model progressively incorporates greater complexity and is posited to capture a broader range of return variations.

Empirical tests were conducted on 25 Portfolios Formed on Size and Book-to-Market, obtained from Kenneth French's website. Monthly returns data span from July 1963 to June 2023, with the first five years earmarked for estimation and the remainder for backtesting. The models were evaluated based on their annualized out-of-sample Root Mean Square Error (RMSE):

Table 5.2: Annualized Out-of-Sample RMSE for Factor Models. This table displays the RMSE of different factor models used in estimating systematic risk across 25 Portfolios formed on Size and Book-to-Market. The comparison is made among the Rolling OLS, Kalman Filter, and the Proposed Model, highlighting their predictive accuracy over the out-of-sample period.

Model	**Rolling OLS**	**Kalman Filter**	**Proposed Model**
Historical Mean	18.54%	16.83%	16.34%
Market Model	7.41%	5.67%	5.91%
FF 3 Factor	1.56%	1.22%	1.32%

Table 5.2 suggests that the proposed model achieves parity with the Kalman Filter and surpasses the Rolling OLS in predictive accuracy, underscoring its potential in systematic risk prediction within financial markets.

5.3 Inflation Risk Premium

The quantification of risk premia of non-traded macro factors stands as a pivotal aspect of modern financial analysis, particularly within the context of asset pricing. This section

delineates the application of the proposed GLS model to estimate the inflation risk premium using an extensive dataset encompassing all U.S. stocks from CRSP, Fama-French factors, and PCE deflator data from January 1963 to December 2022.

The proposed GLS model is particularly adept at estimating the inflation risk premium due to its enhanced ability to handle smaller sample sizes and its robustness in scenarios where the normality of data is not assumed. Unlike the Kalman Filter and Rolling OLS, the GLS model's sophisticated parameter estimation capabilities allow for more accurate and dynamic adjustment to market changes, a crucial factor in accurately capturing the inflation risk premium. This is especially pertinent in the analysis of non-traded macro factors, where traditional models often struggle with the complexity and variability inherent in inflation data. The GLS model's flexibility and precision in handling such nuances make it an invaluable tool in this context, providing deeper insights and more reliable predictions than its counterparts.

Adhering to the Fama and MacBeth 1973 procedure, the first step employs the non-causal GLS filter to estimate time-variant risk loadings via a four-factor model.

$$r_t^i = \begin{bmatrix} r_t^m & \text{SMB}_t & \text{HML}_t & \pi_t & 1 \end{bmatrix} \begin{bmatrix} \beta_t^i \\ s_t^i \\ h_t^i \\ p_t^i \\ \mu_t^i \end{bmatrix} + \varepsilon_t^i \tag{5.8}$$

$$\begin{bmatrix} \beta_t^i \\ s_t^i \\ h_t^i \\ p_t^i \\ \mu_t^i \end{bmatrix} = c + A \begin{bmatrix} \beta_{t-1}^i \\ s_{t-1}^i \\ h_{t-1}^i \\ p_{t-1}^i \\ \mu_{t-1}^i \end{bmatrix} + \nu_t^i \tag{5.9}$$

The subsequent step involves a rolling panel regression correlating returns with these risk loadings, further refined through the non-causal GLS filter. Parameters are iteratively estimated as suggested in Section 4.3.3.

$$r_t^{e,i} - \widehat{\mu}_t^i = \begin{bmatrix} \widehat{\beta}_t^i & \widehat{s}_t^i & \widehat{h}_t^i & \widehat{p}_t^i & 1 \end{bmatrix} \begin{bmatrix} \lambda_t^m \\ \lambda_t^s \\ \lambda_t^h \\ \lambda_t^\pi \\ \alpha_t \end{bmatrix} + e_t^i \tag{5.10}$$

$$\begin{bmatrix} \lambda_t^m \\ \lambda_t^s \\ \lambda_t^h \\ \lambda_t^\pi \\ \alpha_t \end{bmatrix} = d + B \begin{bmatrix} \lambda_{t-1}^m \\ \lambda_{t-1}^s \\ \lambda_{t-1}^h \\ \lambda_{t-1}^\pi \\ \alpha_{t-1} \end{bmatrix} + \xi_t \tag{5.11}$$

Figure 5.1 below illustrates the estimated inflation risk premium over time, with notable peaks corresponding to periods of significant economic turmoil. For example, the first two

peaks around 1974 and 1979 correspond to the two oil crises in the 1970s. The recent surge following the COVID-19 pandemic highlights significant inflationary periods. By capturing key economic events, the figure underscores the model's responsiveness to market dynamics.

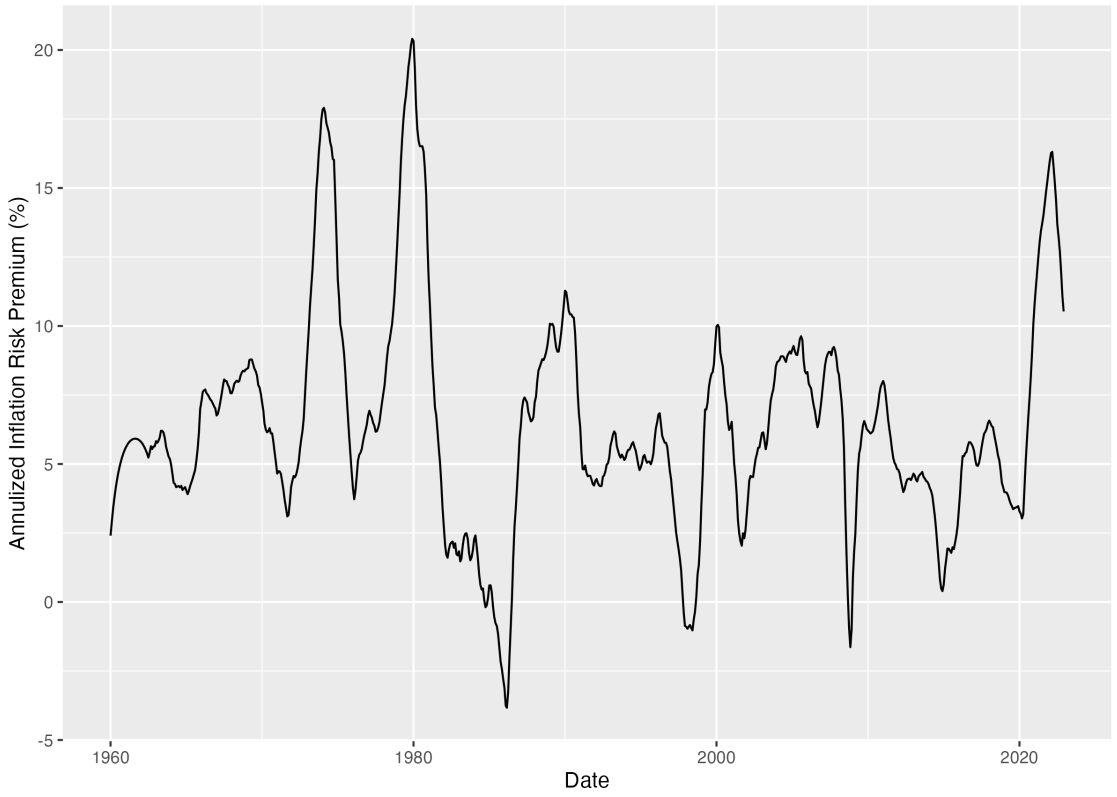


Figure 5.1: Estimated Inflation Risk Premium Over Time. This graph displays the annualized estimated risk premium of inflation, derived from the proposed GLS model. The timeline captures key economic events that influenced inflation expectations, with peaks corresponding to periods of high inflation and economic stress, providing insight into the model's capability to capture inflation risk premium dynamics.

Chapter 6

Conclusions and Further Directions

6.1 Conclusions

In conclusion, this study represents a significant advancement in financial econometrics through the introduction of a novel rolling Generalized Least Squares (GLS) model for dynamic linear models. With varying levels of complexity, four theoretical models are derived, encompassing both causal and non-causal versions to cater to distinct applications. For instance, the one-sided causal version is ideal for forecasting, whereas the two-sided non-causal version is more suitable for retrospective economic and financial research.

In addition, the iterative rolling feasible GLS approach derived herein represents a relevant method for achieving optimal performance with unknown parameters, given the data. What's more, the strategic development of an optimal window size determination methodology further illustrates the thesis's practical utility. This optimization provides a valuable tool for practitioners in the field who prefer using heuristic methods, aiding in the more accurate forecasting of risk and return.

Rigorous simulations show that the proposed strategy stochastically outperforms traditional passive models. When compared to established statistical and econometric models like the Baxter-King Filter and the Kalman Filter, the new method consistently shows superiority. To reiterate, the GLS model's notable strengths in accommodating smaller sample sizes and its resilience against deviations from normality assumptions mark it as a particularly versatile and reliable tool in financial modeling.

This is corroborated by empirical results with real-world data, solidifying the model's universality, robustness, and efficiency. These findings underscore the potential of the rolling GLS model to revolutionize the estimation of time-varying parameters.

6.2 Future Research

Building upon the promising results of the GLS model demonstrated in this study, the next phase of research involves its broader application and a more detailed statistical analysis.

First, a subsequent step of this research will venture into applying the novel rolling GLS model to a wider array of datasets across various domains, such as finance, economics, even robotics and aerospace, thereby establishing its practical utility.

Furthermore, developing a comprehensive statistical inference framework, which includes confidence interval computation and hypothesis testing, is imperative to validate and substantiate the model's estimates rigorously. This framework aims to bridge the gap between theoretical econometric models and their pragmatic deployment in finance.

In summary, these advancements will not only solidify the understanding of the GLS model's efficacy in practical scenarios but also leverage its strengths in handling smaller sample sizes and non-normality to enhance factor loadings estimation and portfolio management strategies. The ultimate aim is to provide a robust framework that enables both financial and non-financial professionals to navigate complex and dynamic environments with greater confidence and precision.

References

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1913610 (visited on 12/04/2023).

Kalman, R. E. (Mar. 1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1, pp. 35–45. ISSN: 0021-9223. DOI: 10.1115/1.3662552. URL: https://doi.org/10.1115/1.3662552.

Rauch, H. E., F. Tung, and C. T. Striebel (1965). "Maximum likelihood estimates of linear dynamic systems". In: *AIAA Journal* 3.8, pp. 1445–1450. DOI: 10.2514/3.3166. URL: https://doi.org/10.2514/3.3166.

Hamilton, James D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". In: *Econometrica* 57.2, pp. 357–384. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1912559 (visited on 11/30/2023).

Doucet, Arnaud and A.M. Johansen (2011). *A Tutorial on Particle Filtering and Smoothing: Fifteen years Later*. 1st ed. Oxford, UK: Oxford University Press.

Julier, Simon J. and Jeffrey K. Uhlmann (1997). "New extension of the Kalman filter to nonlinear systems". In: *Signal Processing, Sensor Fusion, and Target Recognition VI*. Ed. by Ivan Kadar. Vol. 3068. International Society for Optics and Photonics. SPIE, pp. 182–193. DOI: 10.1117/12.280797. URL: https://doi.org/10.1117/12.280797.

Wan, E.A. and R. Van Der Merwe (2000). "The unscented Kalman filter for nonlinear estimation". In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pp. 153–158. DOI: 10.1109/ASSPCC.2000.882463.

Hodrick, Robert J. and Edward C. Prescott (1997). "Postwar U.S. Business Cycles: An Empirical Investigation". In: *Journal of Money, Credit and Banking* 29.1, pp. 1–16. ISSN: 00222879, 15384616. URL: http://www.jstor.org/stable/2953682 (visited on 11/30/2023).

Cogley, Timothy and James M. Nason (1995). "Output Dynamics in Real-Business-Cycle Models". In: *The American Economic Review* 85.3, pp. 492–511. ISSN: 00028282. URL: http://www.jstor.org/stable/2118184 (visited on 12/11/2023).

Baxter, Marianne and Robert G. King (1999). "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series". In: *The Review of Economics and Statistics* 81.4, pp. 575–593. ISSN: 00346535, 15309142. URL: http://www.jstor.org/stable/2646708 (visited on 11/30/2023).

Markowitz, Harry (1952). "Portfolio Selection". In: *The Journal of Finance* 7.1, pp. 77–91. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2975974 (visited on 12/04/2023).

Fama, Eugene F. and James D. MacBeth (1973). "Risk, Return, and Equilibrium: Empirical Tests". In: *Journal of Political Economy* 81.3, pp. 607–636. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1831028 (visited on 12/04/2023).