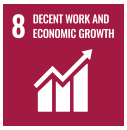


Univariate and Multivariate Inverse Gamma Stochastic Volatility Models

A Dissertation
Submitted to the National Graduate Institute for Policy Studies (GRIPS)
in Partial Fulfilment of the Requirements for the Degree of

Ph.D. in International Economics

by

Blessings Majoni

March 2024

Abstract

This dissertation makes several contributions to the stochastic volatility model literature. The key contribution is that it obtains a novel closed form expression of the likelihood for a stationary inverse gamma Stochastic Volatility (SV) model. As a result, using this expression of the log likelihood, it is possible to obtain the Maximum Likelihood Estimator (MLE) for this class of non linear non gaussian state space models for the univariate model. The dissertation provides two empirical studies to demonstrate this approach.

First, chapter 3 proposes a novel method to explicitly calculate the likelihood for a stationary inverse gamma Stochastic Volatility model, which is conventionally approximated using sampling methods. The derived likelihood is expressed by infinite series of functions and its calculation is implemented by truncating higher order terms. This expression of the likelihood allows the estimation of the parameters and unobserved states for this model class by MLE. Further, the chapter provides the analytical expressions for both the filtering and smoothing distributions of the volatilities as mixture of gammas and therefore it is able to provide the smoothed estimates of the volatility. The chapter shows that by marginalising out the volatilities, the model that is obtained has the resemblance of a GARCH in the sense that the formulas are similar, which simplifies computations significantly. Another significant contribution of this chapter is that, the computer code to perform the analysis has been publicly made available through a R package that is freely available from the Comprehensive R Archive Network (CRAN).

The performance of the proposed method was evaluated using several macroeconomic and financial data sets and its results show that the proposed method achieved accurate calculation of the likelihood with low computational cost. The macroeconomic application uses quarterly inflation data for four countries, that is, UK, USA, Japan and Brazil. A range of other models are also estimated to evaluate the empirical performance of the proposed model. The proposed model performs better than all other models in 50% of the applications in terms of the Bayesian Information Criterion (BIC), with very large gains for the Brazil dataset. The second application uses exchange rates data for 7 currencies (GBP, EUR, JPY, CND, AUD, BRL, ZAR) and finds that the empirical fit of

the proposed model is overall better than alternative models in 4 of the 7 currencies in terms of the BIC, being much superior in the case of currencies with turbulent episodes, such as the Brazilian Real.

Lastly, using daily returns for 8 stocks on the Tokyo Stock Price Index (Topix) and the Standards and Poors 500 ($S\&P500$), the performance of the model is compared to the stochastic volatility model with leverage. The proposed model performs better 62.5% than the asymmetric SV model in terms of the Bayesian Information Criterion (BIC).

Chapter 4 extends the univariate approach to estimating large Vector Autoregressions (VAR) in a multivariate common stochastic volatility (CSV) model. CSV models capture the commonality that is often observed in volatility patterns. However, they assume that all the time variation in volatility is driven by a single multiplicative factor. Given the empirical evidence on fat tailed distributions, and the commonality that is observed in volatility patterns, this model combines stochastic volatility, heavy tailed distributions and a common heteroscedastic latent factor. The volatility for this novel CSV model follows an inverse gamma process, which implies student-t type tails for the observed data. An analytic expression for the likelihood is obtained for this CSV model, which facilitates model comparison. This model is estimated using 4 macroeconomic applications that use 20 variables each for Japan, Brazil, US, and the UK. A second application uses financial data of daily exchange rate returns for a small VAR of 4 currencies and a larger VAR of 8 currencies. The comparison method is based on marginal likelihoods and the one step ahead out of sample predictive likelihoods. The proposed model is compared to other CSV models proposed in the literature. Using two evaluation periods of 24 and 50 data points and two priors, the empirical fit of the common factor inverse gamma SV model has better forecasting accuracy compared to alternative CSV models for the macroeconomic application in 13 of the 16 instances. The BVAR-CSV model is best for the Japan data over a longer evaluation period. Using the alternative prior, over the shorter period, the BVAR-CSV-MA-t is best for the UK while the BVAR-CSV-t model is best for Brazil. In the financial application, the proposed model performs better than alternative models when using the first prior, while the BVAR-CSV-t is best using the alternative prior.

Dedication

Dedicated to my husband Tinashe Samanga and sons Takundiswa, Kansha and Zaine Samanga. To my dear parents Judith and Keveni Majoni and to my brother and sisters.

Acknowledgement

My uttermost gratitude to my immensely supportive supervisor, Roberto León Gonzalez for the insightful guidance, tutorage and constant encouragement throughout my studies. Words alone seem inadequate to express how much he impacted my life and I feel extremely lucky to have been mentored by him. Simply put, the results in this dissertation would have been impossible without his guidance and input. I thank Takenouchi Takashi and Tsuchiya Takashi for the helpful advise and suggestions that helped shape the paper considerably.

Special thanks to Tinashe, my husband for the patience, support and encouragement. I am especially grateful for the assistance in proof reading this thesis even with its seemingly meaningless mathematics and formulas. More-so for listening to countless presentations each time I needed an audience. Finally, I am grateful to my 3 children, Takundiswa, Kansha and Zaine for understanding whenever I failed to attend countless soccer matches, gymnastics and some school events. I am especially grateful for giving me a reason each day to improve myself.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

Stochastic volatility models by definition, are models in which the volatility itself is stochastic. These models were first proposed by Black (1976) for financial econometrics where they were found to have better fit to volatility that is observed in historical equity returns and modelling the leverage effect between the returns and volatility. Early applications of stochastic volatility models for macroeconomics, using Bayesian methods were proposed by Jacquier et al. (1994) whereby estimation of the algorithm was achieved by sampling the volatilities individually.

Literature is replete with a wide class of varying stochastic volatility models as researchers attempt to incorporate the desirable properties of traditional models and the properties observed in macroeconomic and financial variables. One of the early influential and efficient formulations of stochastic volatility proposed by Kim et al. (1998) is represented as follows:

$$y_t = e_t \exp\left(\frac{h_t}{2}\right)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \delta_t \sigma$$

where y_t represents the observed data, e_t is a multiplicative stochastic error , h_t is the log of the volatility which is a latent factor, μ is the mean of the log of volatility. ϕ in this model is a parameter for the persistence of the term for the volatility. e_t and δ_t are both gaussian with mean 0 and unit variance.

The initial distribution of the latent parameter has a Gaussian distribution as follows:

$$h_1 \backsim \text{normal}\left(\mu, \frac{\sigma}{\sqrt{1-\phi^2}}\right)$$

Kim et al. (1998), showed that the log volatilities of stochastic volatility models can be marginalised out to estimate the unknown parameters. Their approach is based on a

Markov Chain Monte Carlo (MCMC) method and is known in the literature to be efficient and reliable and has been used liberally in both macroeconomic and financial applications. Such an approach follows a log normal specification of the volatilities which implies that the distributions have thin tails (Madan & Seneta, 1990). However, a key property of macroeconomic and financial data is that they exhibit fat tails in their distribution. Thus, one drawback of the above approach to stochastic volatility specifications is the inability to capture the fat tails. In addition, this model does not allow for serial dependence in the errors of the measurement equations. However, time series variables are known to have serially correlated errors, a phenomenon called leverage effects in this field.

Succeeding studies to this highly efficient baseline approach to stochastic volatility model estimation attempt to further capture other intrinsic properties of the data while ensuring tractability of the model at the same time. Omori et al. (2007) extends the above approach by providing an approach that can handle stochastic volatility models with leverage effects. Regarding the log normal stochastic volatility model, the model proposed by Omori et al. (2007) has a similar structure to that of the baseline approach by Kim et al. (1998) above, where in this case e_t and δ_t are both gaussian as follows:

$$\begin{pmatrix} e_t \\ \delta_t \end{pmatrix} \backsim \mathrm{N}(0, \Sigma)$$

where:

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}$$

The leverage effect is thus captured by the parameter ρ for this class of non linear state space models. Yu (2005) and Wang et al. (2011), among other studies, provide detailed discussions of alternative approaches that model the leverage effect in stochastic volatility models, with the latter adding fat tailed distributions to this baseline model. Stochastic volatility models continue to capture the interest of researchers because they tend to have better forecasting accuracy than other families of state space models. However, they are known to be challenging to estimate.

In linear gaussian state space models, the likelihood is available in closed form and

can easily be calculated with the Kalman Filter algorithm (e.g. Durbin and Koopman (2012)). However, few studies have attempted to obtain a closed form expression for the likelihood in nonlinear non-Gaussian state space models. In the univariate case, for example, Shephard (1994) obtains a closed form expression for the likelihood of a non-stationary SV model known as Local Scale Model. Uhlig (1997) builds on and generalizes Shephard (1994) to the multivariate case. They obtain an analytic expression for the likelihood and posterior density of a SV non-stationary restricted singular Wishart model. Creal (2017) obtains an analytic expression for the likelihood in a SV gamma model and shows that analytic expressions for the likelihood could also be obtained for a family of non linear non Gaussian state space models. The gamma SV model in Creal (2017) implies a variance-gamma distribution for the data and this distribution has thin tails (Madan & Seneta, 1990).

On account of the above, this dissertation adds to the literature by obtaining a closed form expression of the likelihood for inverse gamma stochastic volatility models. In the absence of a closed form expression of the likelihood, maximum likelihood estimation can not be used to obtain the parameters of the model. In addition, the smoothed estimates of the volatility can not be obtained. The approach to obtaining the likelihood in this dissertation is used to analyse a univariate inverse gamma stochastic volatility model and a multivariate approximate factor model resulting in the two studies covered in this dissertation.

First, chapter 3 considers the univariate inverse gamma stochastic volatility model and obtains the exact likelihood for this model. This expression of the likelihood allows the estimation of the parameters and unobserved states for this non linear non gaussian model class by Maximum Likelihood estimation. Further, it provides the analytical expressions for both the filtering and smoothing distributions of the volatilities as mixture of gammas and therefore it can provide the smoothed estimates of the volatility. The chapter shows that by marginalising out the volatilities, the model that is obtained has the resemblance of a GARCH in the sense that the formulas are similar, which simplifies computations significantly.

The chapter provides two empirical applications using financial and macroeconomic variables. The macroeconomic application uses quarterly inflation data for four countries,

that is, UK, USA, Japan and Brazil. It examines the accuracy of this novel approach by comparing the value of the log likelihood obtained using the value of the parameters at the maximum to the log likelihood obtained using particle filters. A range of other models are also estimated to evaluate the empirical performance of the proposed model. The proposed model performs better than all other models in 50% of the applications in terms of the Bayesian Information Criterion, with very large gains for the Brazil dataset. The second application uses exchange rates data for 7 currencies (GBP, EUR, JPY, CND, AUD, BRL, ZAR) and finds that the empirical fit of the proposed model is overall better than alternative models in 4 of the 7 currencies in terms of the BIC, being much superior in the case of currencies with turbulent episodes, such as the Brazilian Real.

Chapter 4 extends the univariate approach to estimating large Vector Autoregressions (VAR) in a multivariate stochastic volatility model. Given the increase in fat tailed events over the years and the empirical evidence on the co movement that is observed in volatility patterns of macroeconomic data, this model combines stochastic volatility, heavy tailed distributions and a common latent factor. The common latent factor is taken as multiplicative. The chapter provides 2 applications to estimate this model. The first application uses 20 macroeconomic variables each for Japan, Brazil, US, and the UK. The second application uses financial data of daily exchange rate returns for a small VAR of 4 currencies and a larger VAR of 8 currencies. The comparison method uses marginal likelihoods and the one step ahead out of sample predictive likelihoods. The proposed model is compared to other common stochastic volatility models and a factor stochastic volatility model. The empirical fit of the common factor inverse gamma SV model has the better forecasting accuracy compared to other common stochastic volatility models for both the macro economic and financial application. However, compared to the factor stochastic volatility model, it is best at a cut point of 24 for all four macro applications. Increasing the evaluation period, it has a better fit only for Japan and is second best to the factor model overall.

The rest of the dissertation is structured as follows: chapter 2 provides an overview of the study area with particular emphasis on economic factors that shape the data used in the study. Chapter 3 covers the univariate case of the approach to estimating stochastic volatility models using the closed form expressions of the likelihood that we

obtain. Chapter 4 provides the multivariate approach using a CSV inverse gamma model. Chapter 5 provides the R code and tutorial that is published on the CRAN repository, and chapter 6 concludes and provides the policy implications for policy makers.

CHAPTER 2
OVERVIEW OF THE STUDY AREA

2.1 Introduction

The macroeconomic empirical applications for both the univariate and the multivariate chapters, are based on developed high income countries, that is, the US, UK and Japan, and an emerging upper-middle income developing economy Brazil. These economic classifications are based on the World Bank income classifications (World Bank, 2022). The financial application for the univariate model in chapter 3 is based on exchange rates for a mix of developed and developing countries, that is, the EURO (EUR), the Great British Pound sterling (GBP), Japanese yen (JPY), the Australian dollar (AUD), the Canadian dollar (CAD), the South African Rand (ZAR) and the Brazilian Real. The exchange rates application in Chapter 4 is based on Zimbabwe's major trading partner countries for both exports and imports for the year 2020. This list can be derived from the World Integrated Trade Solution (WITS) section of the World Bank data series. As such, the currencies considered are the EUR, GBP, ZAR, Chinese Yuan (CNY), Singapore Dollar (SGD), Hongkong Dollar (HKD) and Indian Rupee (INR). This chapter provides an overview of the monetary policies, exchange rate regimes and foreign exchange rate interventions of all the economies discussed in the empirical applications for the period under study.

2.2 Overview of the Exchange Rate Policies and Interventions

2.2.1 South African Rand

The South African Reserve Bank maintains a floating exchange rate regime. A floating exchange rate regime is whereby a currency's exchange rate is determined by the market forces of supply and demand of other currencies in the international money market. The Reserve Bank occasionally intervenes and smoothes out exchange rate volatility to ensure that the South African Rand is stable without necessarily targeting a specific exchange rate. A key factor that affects floating exchange rates is the inflation rate. As a result, the South African Reserve Bank uses inflation targeting to ensure that the South African Rand is maintained relative to the domestic consumer price inflation index target. For the period under study, that is, from 2019 to 2023, this consumer price inflation target was set to between 3% and 6% (South African Reserve Bank, 2021). Exchange rates typically play a very important role in monetary policy and in some instances can be used as a monetary policy tool.

The major interventions to influence the exchange rate by the South African Reserve Bank, during the period covered by this study were undertaken during the Covid-19 crisis. In March 2020, the Reserve Bank used a number of instruments such as the amendment of the end of day supplementary repo which was changed to the intra-day overnight supplementary repos to inject liquidity into the money markets. The supplementary repos are repurchase operations agreements by the Reserve Bank at a fixed repo rate. In addition, a long term repo was introduced for up to 12 months.

2.2.2 Pound Sterling

Similar to the South African Rand, the pound sterling is a fiat currency whose exchange rate is determined by the market forces of supply and demand. The Bank of England uses monetary policy tools such as inflation targeting to intervene in the exchange rate financial markets. The inflation target for the UK is set to 2%. To ensure alignment to

the inflation target, the Bank of England uses mainly 2 tools. The key tool is the Bank rate, which is the UK's official interest rate. In March 2020, the Bank of England cut the Bank rate to 0.1%. Besides the Bank Rate, the Bank of England in addition uses another monetary policy tool termed Quantitative easing. Quantitative easing mainly involves injecting money directly into the money markets by purchasing of government bonds as well as purchasing other financial assets by a central bank. The amount of these bonds are usually predetermined. The mechanism involved in buying these financial assets result in the prices of these assets and bonds going up, and thus in the long term, interest rates decrease. Lower interest rates inturn result in increase in an increase in household consumption expenditure which in turn increases inflation. During the period March 2019 to March 2023, the bank increased the Quantitative easing amount in March, June and lastly November 2020 (Statista Research Department, 2022). Since November 2022, the Bank of England began reversing the Quantitative easing by selling the bonds. Reversing Quantitative easing is referred to as Quantitative tightening. This has the reverse effect of decreasing inflation rate and increasing the interest rate.

2.2.3 Japanese Yen

The Bank of Japan's exchange rate policy for the Japanese Yen is perhaps closely defined by a "free" floating exchange rate regime. A free floating exchange rate regime is one whereby the monetary authorities do not intervene in the international money markets and allow the exchange rate of the currency to be determined solely by market forces of supply and demand of different foreign currencies. For instance, prior to September 2022, the last intervention by the Japanese monetary authorities, to support the Yen were done as far back as June 1988 (Jun Saito, 2022). The recent $Covid-19$ crisis however, resulted in a rapid increase in the depreciation rate of the Japanese Yen since March 2022. The Bank of Japan as a result intervened in September and October 2022 by drawing down and selling the foreign reserve stocks of the United States Dollar and buying Japanese Yen from the money market amounting to 2.8382 trillion yen in September and 6.3499 trillion yen in October 2022 (Jun Saito, 2022). Such an intervention results in decreasing the amount of Japanese yen in the money market and thus strengthening the yen value

against the US dollar by increasing demand for the yen. The immediate effect is a decrease in the monetary base as deposits held by the private sector shift to the government. A decrease in the monetary balance translates to a decrease in the current account balance in the absence of counter measures to maintain the monetary base level.

2.2.4 Canadian Dollar

In close comparison to the Bank of Japan, the Bank of Canada manages a floating exchange rate regime and is averse to intervening and influencing the exchange rate for the Canadian dollar in the money markets. Foreign exchange interventions in the money market are reserved for extreme volatility in the exchange rate and market breakdown in price determination. Intervention methods for the Bank of Canada are mainly set to involve using the exchange fund account which holds foreign reserves for the government of the top 4 traded currencies, that is, the US dollar, Japanese yen, the pound sterling and euro. The Bank of Canada last intervened to manipulate the exchange rate markets in September 1998. Even though the bank does not intervene, the monetary policy objective for Canada is largely influenced by inflation targeting, thus, exchange rate movements were closely monitored to maintain the rate of inflation at a 2% for the period covered by the study.

2.2.5 Australian Dollar

The Reserve Bank of Australia in similar fashion does not target any specific rate of exchange for its Australian Dollar. It manages a floating exchange rate policy and in parallel to the previous economies, the monetary policy in Australia is conducted under the framework for inflation targeting. Typically intervention transactions in the foreign exchange markets by the Reserve Bank of Australia are in the spot market. The interventions consist of buying and or selling the Australian dollars in the market to shift ownership of the currency from and or to the government.

The Australian Reserve Bank typically uses foreign exchange swaps to sanitize and balance out the effect of these foreign exchange market interventions. During the period under study the Australian Reserve Bank made a number of interventions that directly

affect the exchange rate. In March and November 2020, the cash rate target was reduced to 0.25% and 0.1% respectively. The cash rate refers to the rate of interest paid in the money market by banks when they borrow funds overnight from other banks. Lowering the cash rate translates to lower interest rates, thereby reducing the cost of borrowing for households and the private sector and ultimately reducing the exchange rate. The cash rate was then increased back to normal levels in May 2022. In November 2022, the Australian Reserve Bank also announced a government bonds purchase program for 100 billion bonds. These bond purchases were completed by April 2021. Subsequent similar valued bond purchase programs were announced from February 2021 and additional weekly bonds of 4 billion dollars until February 2022. The effect of such a bond purchase program ia an increase in money holding in the market, lower interest rates, which has a similar effect as the cash rate reduction of lowering cost of borrowing and lowering the exchange rate in favour of the Australian dollar.

2.2.6 Brazilian Real

The Central Bank of Brazil operates a floating exchange rate regime in line with the inflation targeting monetary policy framework. Over the years the Central Bank has been known to intervene frequently in the money market. Interventions in the foreign exchange market are mostly carried out through forex swap auctions, typically to counter excess volatility in the market operations. The forex swaps are currency combination swap contracts consisting of both the spot and forward transactions. For example, the Central Bank of Brazil may buy forex swap contracts in the spot market at a premium to reduce the amount of the foreign currency in the market and smooth out depreciation of the foreign currency exchange rate, and later sell the currency in the forward market. The effect on the money market is similar to the Central Bank selling future sales of the exchange currency which is typically the US dollar. When the bank sells the forex swap contracts, this is termed a reverse forex swap. Such a sale has the reverse effect of decreasing the exchange rate and smoothing out appreciation of the local currency.

During the period under study, the Brazilian Central Bank was arguably, of all the currencies under study, the more volatile as the Brazilian Real depreciated its value

by over 30% at its lowest. As expected during this period the bank made numerous interventions. For example, the bank increased the forex swaps stock to smooth out volatility from the initial sharp depreciation of the Brazilian Real in the early onset of Covid in 2020. In addition, the selic rate, which is the base interest rate for Brazil was reduced to 2% in 2020. However, due to other macroeconomic factors and inflationary shocks, the bank anticipated effect of reducing the selic rate was contrary to the measure as instead, the Real depreciated even further. In response and to smooth out inflation price inflation and devaluation of the local currency, the Brazilian Central Bank increased the selic rate by 0.75%, and subsequent increases resulted in the selic rate level of 4.25% by June 2021. Therefore, in 2021 the volatility of the Brazilian Real was particularly high as can be seen on plots of the currency in Chapter 3. Furthermore, in November 2020 the Central Bank also announced the rolling over of $11.8 billion worth of foreign exchange swap contracts with a maturity of January 2021. In September, and October 2020, the Central Bank sold in unscheduled auctions of $500 million and $1 billion foreign exchange swap contracts.

2.2.7 Euro

The Euro is part of a currency union of 19 member states of the 27 European Union countries. While for the member states that are part of the Euro, the exchange rate policy is a hard peg, with other member states, the Euro regime is a floating exchange rate against other currencies. A hard peg exchange rate is whereby the monetary authorities target a specific exchange rate against other countries exchange rates. Involvement of the monetary authority in determining the exchange rate instead of allowing the economic forces of supply and demand to determine the rate is what distinguishes this regime from the floating regime.

The European Central Bank manages the monetary policy for the Euro Zone. The bank intervenes by using monetary policy tools such as interest rates and quantitative easing. Interventions are done either as coordinated interventions where the bank intervenes with the assistance of other member state's central banks, or sole interventions by the bank itself in unilateral interventions. Prior to the period under study, the bank had

only intervened in the foreign exchange rate market at inception in the year 2000 and in 2011. In response to the coronavirus pandemic, a major intervention by the European Central Bank was the introduction of a pandemic emergency purchase program in March 2020 to safeguard the economy of the Euro area and to stabilise the money market. The implementation of this program resulted in a purchase of securities amounting to $1.7 trillion between March 2020 and March 2022 from both the private and public sector in the Euro area.

2.3 Overview of Monetary Policies

The macroeconomic applications in Chapter 3 and 4 are all based on 4 countries, that is, the US, UK, Japan and Brazil. The applications for the first 3 countries cover the period 1960 to 2023, whereas for Brazil cover the period 1994 to 2023 due to unavailability of historical observations. The main aspects of monetary policy strategies for the developed countries above are largely similar, as such, the variables under study are expected in the long run to be affected in a similar way by economic shocks all things being equal.

The main objective in their monetary policy strategy implementation for the Bank of Japan, Federal Reserve of the US, and the Bank of England is price stability of both goods and services through an inflation targeting framework. The target is set at 2% of the inflation rate for all these countries. The Federal Reserve Bank is in charge of three monetary policy tools, that is, the discount rate, open market operations and reserve requirements. Using these tools the Federal Reserve can control the federal funds rate, money supply and demand and deposits held by financial institutions. Thus monetary policy ultimately affects the behaviour of variables in the economy.

The Bank of England uses quantitative easing explained above and the bank rate, which is the rate at which banks are charged when borrowing money from the central bank. Similar to the Federal Reserve, the Bank of Japan uses open market operations, reserve requirements and the call rate as tools of monetary policy. The Central Bank of Brazil uses the minimum reserve requirement, open market operations, and the discount rate to influence the selic rate which is the main monetary policy tool. The monetary policy framework follows an inflation targeting framework, with an inflation target of

3.2% for 2023 which is higher in comparison to the 3 developed countries above. Communication and transparency is another key instrument used in this inflation targeting framework by the Central Bank of Brazil.

CHAPTER 3
EXACT LIKELIHOOD FOR INVERSE GAMMA STOCHASTIC VOLATILITY MODELS

3.1 Introduction

For most non-linear or non-Gaussian state space models it is difficult to obtain the likelihood function in closed form. This prevents the use of Maximum Likelihood Estimation (MLE). As a result most studies use Bayesian estimation with Markov Chain Monte Carlo (MCMC) methods. Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models are simpler to estimate than Stochastic Volatility (SV) models, because the likelihood for a GARCH model can be easily calculated in closed form (e.g. Engle (1982), Bollerslev (1987)). However, SV models have often been found to outperform GARCH models in empirical studies for both macroeconomic and financial data (e.g. Chan & Grant (2016a) and Kim et al. (1998)). In addition, unlike GARCH models, SV models provide not only filtered estimates but also smoothed estimates of the volatility.

Although in linear Gaussian state space models the likelihood is available in closed form and can easily be calculated with the Kalman Filter algorithm (e.g. Durbin and Koopman (2012)), few studies have attempted to obtain a closed form expression for the likelihood in nonlinear non-Gaussian state space models. Shephard (1994) obtains a closed form expression for the likelihood of a non-stationary SV model known as Local Scale Model, showing the similarities to GARCH models. Uhlig (1997), builds on and generalizes Shephard (1994) to the multivariate case. They obtain an analytic expression for the likelihood and posterior density of a SV non-stationary restricted singular Wishart model. Creal (2017) obtains an analytic expression for the likelihood in a SV gamma model and shows that analytic expressions for the likelihood could also be obtained for a family of non linear non Gaussian state space models. The gamma SV model in

Creal (2017) implies a variance-gamma distribution for the data and this distribution has thin tails (Madan & Seneta, 1990). In contrast, inverse gamma SV models imply a student-t distribution, thus, they can account for the fat tails that are observed in most macroeconomic and financial data (Leon-Gonzalez, 2018).

The purpose of this study is to obtain an analytic expression of the likelihood for the inverse gamma SV model. This exact likelihood solution will allow the estimation of the parameters and unobserved states for this non linear and non gaussian state space model by MLE. Without the likelihood expression, estimation of non linear non gaussian state space models generally involves bayesian methods such as Markov Chain Monte Carlo. We show that by marginalising out the volatilities, the model that we obtain has the resemblance of a GARCH in the sense that the formulas that we get are similar, which simplifies computations significantly. Moreover, the likelihood function proposed in this paper can be calculated efficiently using a simple recursion. The calculations can be accelerated by doing computations in parallel, as well as by applying Euler or other acceleration techniques to the Gauss hypergeometric functions in the likelihood. In addition to obtaining the exact likelihood, we obtain analytically the expressions for the smoothed and filtered estimates of the volatilities. We provide the computer code to calculate the likelihood as a user-friendly R package that is available on CRAN.

Section 3.2 reviews the literature on previous attempts to obtain analytically the likelihood expressions for non linear non gaussian state space models. Section 3.3 describes our model and derives the analytic expression of the likelihood. In addition, the section provides the analytic expressions for the filtering and smoothing distributions of the volatilities. Section 3.4 evaluates the empirical performance and computational efficiency of the proposed novel algorithm with a comparison to other methods. We provide empirical applications using exchange rates data for 7 currencies to the US dollar and quarterly inflation data for four countries. Section 3.5 concludes.

3.2 Literature Review

3.2.1 Stochastic Volatility Models with an Exact Likelihood

There are very few non linear non gaussian state space models for which the likelihood can be obtained exactly. In what follows we review some of the SV models for which an analytic expression of the likelihood has been obtained.

To obtain the maximum likelihood estimates for a generalised non stationary local scale model, Shephard (1994) uses the conjugacy between the gamma and the beta distribution. Using our notation, their model for a univariate observed variable y_t can be expressed as:

$$y_t = x_t\beta + h_t^{-\frac{1}{2}} e_t, \qquad\qquad e_t \sim N(0,1)$$

where x_t is a vector of predetermined variables which could include lags of y_t, and the inverse of h_t is the time varying volatility. The law of motion for the volatilities is:

$$h_{t+1} = h_t \frac{\nu_t}{\lambda} \qquad\qquad \nu_t \sim Beta(\alpha_1, \alpha_2) \tag{3.2.1}$$

with $\alpha_2 = \frac{1}{2}$. The initial distribution is a gamma with parameters ν and S_1 such that h_1 has the following density function:

$$f(h_1|S_1) = h_1^{\frac{\nu}{2}-1} \exp\left(-\frac{h_1}{2S_1}\right) \frac{1}{\Gamma(\nu/2)(2S_1)^{\frac{\nu}{2}}} \tag{3.2.2}$$

where for mathematical convenience the initial density is restricted such that $\alpha_1 = \frac{\nu}{2}$. The parameters to be estimated are β, ν, λ and S_1. Note that, in contrast with the other models in this paper, the volatility follows a non-stationary process. As shown in Appendix A.1.6, defining $Z = h_1 - \lambda h_2$ for $\in (0, \infty)$, the likelihood for this model can be obtained by integrating over the state variable Z. Given that the process for the stochastic volatility is multiplicative, the likelihood is as follows:

$$\pi(y_t|y_{1:t-1}) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)} \lambda^{\alpha_1} \left(\frac{S_{t+1}}{S_t}\right)^{\alpha_1} \frac{1}{\sqrt{2\pi}} \left(2\left((y_t - x_t\beta)^2 + \frac{1}{S_t}\right)^{-1}\right)^{\alpha_2} \tag{3.2.3}$$

where $S_t = \left((y_{t-1} - x_{t-1}\beta)^2 + \frac{1}{S_{t-1}}\right)^{-1}\frac{1}{\lambda}$ and $y_{1:t-1} = (y_1, y_2, ..., y_{t-1})$. To facilitate the reading here and in the following we do not write explicitly x_t as a conditioning argument.

The framework in Shephard (1994) provides a formal justification to Bayesian methods of variance discounting used in earlier literature (West & Harrison (2006), p.p. 360-361).

Creal (2017) shows that closed form solutions for the likelihood can be obtained for a family of non linear state space models with observation densities $p(y_t|h_t, x_t; \theta)$, in which the continuous valued time varying state variable h_t can be analytically integrated out conditionally on a discrete auxiliary variable z_t. x_t in these models are the predetermined regressors and θ is a parameter vector. The models in this class are defined as follows:

$$
\begin{aligned}
y_t &\backsim p(y_t|h_t, x_t; \theta) \\
h_t &\backsim Gamma(\nu + z_t, c) \\
z_t &\backsim Poisson\left(\frac{\phi h_{t-1}}{c}\right)
\end{aligned}
$$

where c is a scale parameter and ϕ determines the persistence of the state variable. For example Creal (2017) provides the following two alternative sufficient conditions for being able to integrate analytically these densities conditional on z_t:

$$
\begin{aligned}
p(h_t|\alpha_1, \alpha_2, \alpha_3) &\propto h_t^{\alpha_1} \exp(\alpha_2 h_t + \alpha_3 h_t^{-1}) \\
p(h_t|\alpha_1, \alpha_2, \alpha_3) &\propto h_t^{\alpha_1}(1 + h_t)^{\alpha_2} \exp(\alpha_3 h_t)
\end{aligned}
$$

where $\alpha_{1:3}$ are functions of only the observations and parameters of the model. Thus, the contribution to the likelihood of one observation conditional on z_t can be obtained by integrating out the continuous state variables h_t analytically. The model that is obtained after integration simplifies to a Markov Switching model over the support of the non-negative discrete state variables z_t. The likelihood for these Markov Switching models can therefore be obtained recursively. Creal (2017) gives the detailed recursive formulas to obtain the likelihood for some specific models within this family, such as the gamma stochastic volatility models, stochastic duration models, stochastic count models and cox processes.

The gamma SV model by Creal (2017) can be expressed as follows:

$$y_t = \mu + x_t\beta + \gamma h_t + \sqrt{h_t}e_t, \qquad e_t \sim N(0,1)$$

where γ determines the skewness. When $\gamma = 0$ the model implies a variance-gamma distribution for the observed variable, which has thin tails (Madan & Seneta, 1990). The initial stationary distribution is $h_1 \backsim \text{Gamma}\big(\nu, \frac{c}{1-\phi}\big)$ and the unconditional mean is $E(h_1) = \frac{\nu c}{1-\phi}$.

More recently, Sundararajan & Barreto-Souza (2023) propose a composite likelihood approach for the same model that we analyze in this paper, and which was estimated with Bayesian methods earlier by Leon-Gonzalez (2018). While they do not obtain the MLE as we do, their approach uses an expectation maximization algorithm to find the maximum of the composite likelihood, albeit with some restrictions.

3.2.2 Bayesian Evaluation of Forecasting and Model Comparison Methods

The basis of comparison used in this paper are log likelihoods and the Bayesian Information Criterion (BIC). When the number of parameters are the same, a larger value of the likelihood typically implies better forecasts. The BIC sums the log likelihood with a penalty term that accounts for model complexity as follows:

$$BIC = -2 \times \text{log likelihood} + d \times (log(n)) \tag{3.2.4}$$

where n is the sample size, d is the dimension of the estimated parameter space. Written in this way, the BIC is thus $\frac{1}{2}$ times the Schwartz Criterion (Schwarz, 1978).

The BIC puts more importance on forecasts and does so by penalising for the complexity of a model imposed using the term $d \times (log(n))$ to reduce and mitigate for over-fitted models (Bishop, 2007). Thus, the BIC favours models with less parameters and of lower dimensions (Schwarz, 1978).

An alternative Information criterion that is equally useful for forecasting is the Akaike Information Criterion (AIC). The AIC does something similar to the BIC given the

relation to the likelihood, and it is obtained as:

$$AIC = -2 \times \log \text{ likelihood} + 2 \times d$$

Akaike (1974). Thus, the BIC and AIC differ as a result of the penalty term. The asymptotic property for the BIC is that of consistency (as proved in the papers cited above). Shao (1997) proves for linear models that the AIC is asymptotically the same as cross validation which is a method that is used to estimate the prediction error [1]. The AIC minimises the expected Kullback-Leibler divergence as a result. Shibata (1981) similarly proves the asymptotic optimality property of the AIC for the non-parametric regression case where the dimension is infinite, where he used the AIC as a selection rule to determine the number of terms required to be retained in a Fourier series expansion. This optimality property means that supposing all of the models are false, provided that the number of similar sized models does not grow exponentially, the AIC tends to result in choosing the model that is the closest in terms of similarity to the true model see e.g Shibata (1983) and Shao (1997).

Supposing instead that a true model exists, then the BIC tends to select the true model with probability approaching to 1 as the sample size n goes to infinity as proved in e.g Nishii (1984) and Schwarz (1978). In this case the AIC would instead tend to choose more complex models asymptotically. However, for finite samples, because of the heavier penalty for complexity of models imposed by the BIC compared to the AIC, the former is known to often choose too simple models. Thus, the choice of BIC or AIC for model selection is not so clear (Hastie et al., 2009). Generally, the BIC is known to be more appropriate for explanatory modelling whereas the AIC tends to be preferred for prediction(i.e to measure predictive accuracy). All these considerations will thus have an impact on the forecasts. Sakamoto et al. (1986) provide a detailed discussion on the applications of the AIC. Arguments in their analysis point out that, given the formulations and comparing the two criterion together, the AIC tends to be less accurate in the case where the number of parameters of the model are too many, or the actual and true model is small. The AIC is connected to statistical modelling as shown by (Konishi

[1] Stone (1977) also proves albeit under some weak conditions, the asymptotic equivalence of the AIC to the Leave One Out cross validation

& Kitagawa, 2008).

The BIC is an approximation to the marginal likelihood asymptotically. Schwarz (1978) proves this property for the family of observations with a density that come from the Koopman-Darmois family [2]. Given the BIC formula in (3.2.4) and given that the marginal likelihood denoted as $p(y)$, can be conveniently written as:

$$p(y) = \int_{\Theta} p(y|\theta)\pi(\theta)d\theta$$

where $p(y|\theta)$ is proportional to the likelihood for the model and $\pi(\theta)$ is the prior density, the marginal likelihood relates to the BIC in that minus 2×log marginal likelihood approaches the BIC value asymptotically. Since the parameters θ are marginalised out, the marginal likelihood is sometimes known as the integrated likelihood. In addition, the marginal likelihood in Bayesian analysis is a normalising constant given that it allows the posterior density to integrate to 1.

The marginal likelihood itself is an accumulated evaluation of out of sample forecasts like the predictive likelihood. Predictive likelihoods are an alternative method for model evaluation and are used to evaluate density forecasts. Predictive likelihoods do this by evaluating the predictive density for a model M at some horizon h *ex ante*, evaluated at the actual value *ex post*. The predictive likelihood is larger the more likely it is for the actual outcome to be realised given the density forecast. Thus, given the parameter vector θ_M, conditional on the previous data y_{t-1} and the model under comparison M, the one step ahead predictive likelihood can be evaluated as follows:

$$p(y_t^{\circ}|y_{t-1}^{\circ}, M) = \int_{\Theta_M} p(y_t^{\circ}|y_{t-1}^{\circ}, \theta_M, M)p(\theta_M|y_{t-1}^{\circ}, M)d\theta_M$$

A compelling merit of predictive likelihoods is that they are not very sensitive to the prior distribution of the density, Geweke & Amisano (2010) has a more complete discussion on this and shows how this integral can be approximated.

The marginal likelihood can be construed as the product of one step ahead predictive likelihoods as follows:

$$p(y_T^{\circ}|M) = \prod_{t=1}^{T} p(y_t^{\circ}|y_{t-1}^{\circ}, M)$$

[2] see also Murphy (2012) for a more recent detailed explanation

Geweke & Amisano (2010) provide a more detailed analysis on how the predictive likelihood is connected to the marginal likelihood [3]. Another method that is similar to the log of the predictive likelihood is the log score.

Good (1952) initially proposed the log scoring rule defined as the log predictive density of an outcome y_{t+k} that is evaluated at the realised outcome y°_{t+k}. Predictive density forecasts can be evaluated based on the average logarithmic score (Diebold & Lopez, 1996). The log score for a model is defined as follows:

$$\text{Log Score}(y^{\circ}_t, M) = \sum_{t=1}^{T} \log(p(y^{\circ}_t; y^{\circ}_{t-1}, M))$$

as in Geweke & Amisano (2011), where $p(y^{\circ}_t; y^{\circ}_{t-1}, M)$ is the predictive density. Bernardo (1979) demonstrates for the continuous case of $\{y_t\}$ that given the scoring rule is based on its predictive density value at the realised outcome y°_t, a key property of the log score is that it is 'the only proper local score'[4].

A score is called proper where it satisfies some necessary conditions. For example, the log score penalises heavily those density forecasts that would have resulted in an actual outcome being assigned very low probability, that is , where y°_t is located in a region of low predictive density (Gneiting & Raftery, 2007), (Mitchell & Hall, 2005), (Amisano & Giacomini, 2007) and (Winkler & Murphy, 1968). In addition, the score assigns high scores where the actual outcome is assigned a high probability and thus is located in a region of high predictive density.

This 'the only proper local score' is as described in Geweke & Amisano (2011) in the following form:

$$g(y^{\circ}_t) + c \sum_{t=1}^{T} \log(p(y^{\circ}_t; y^{\circ}_{t-1}, M))$$

where $c > 0$. Thus, for the discrete case, the proper local scoring rule is just a linear transformation of the log score defined above over the support of at least three discrete points of a finite set of $\{y_t\}$ as proved by Shuford Jr et al. (1966)[5].

[3]For a discussion on how the likelihood is linked to the predictive density see Geweke (2005)

[4]Gneiting & Raftery (2007, p.366) also discusses this continuous case in detail

[5]See Geweke & Amisano (2011), Winkler (1969), DeGroot & Fienberg (1981) and Clemen et al. (1995) for additional discussion of the discrete outcome

3.2.3 Why Forecasting is Important for Policy

Policy makers rely on forecasts to assess the impact of reforms ex ante. The role played by forecasting on policy making decisions cuts across various policy making decisions such as seasonal climate forecasting (e.g Lemos et al. (2002)), or macroeconomic forecasting, (e.g Leon-Gonzalez et al. (2021)). Central banks for example largely depend on forecasting because of the time lag in the transition mechanisms of monetary policy. As a result of this lag, macroeconomic variables such as inflation, have to be forward looking as they take time to respond to this mechanisms.

The importance of the role played by forecasting on policy dates back to the pioneering works by Tinbergen (n.d.) and Theil (1958). These pioneering works provided practical examples on the role required to be played by forecasting in policy areas such as employment policy. Given how broad the role of forecasts is in policy making, we narrow the review on this topic by providing a few high impact examples.

A good practical example of using forecasting in policy making is the 'Fan charts' produced by central banks such as the Bank of England see e.g Bank of England (2022), the Reserve Bank of Australia see e.g David, Reifschneider and Peter, Tulip (2017) and the Federal Reserve of the United States see e.g the announcement in Par (2006). The use of these Fan charts was introduced by the Bank of England in 1996 mainly as a means of communicating publicly their inflation forecasts in terms of potential implications and uncertainty Britton et al. (n.d.). The Fan charts such as those used by the Bank of England use predictive density to illustrate and project where inflation levels could be and measure forecast uncertainty ((Mitchell & Hall, 2005) and Groen et al. (2009)). The Fan charts are currently provided for many other variables such as GDP growth projections and not just inflation following the literature advocating for the release of more forecast information to the public e.g Svensson (1997)'s argument for including output forecasts. On the other hand, a diverging literature such as in Mishkin (2004) asserted against providing the forecasts for some variables such as output gap or the objective function of a bank given the risk of misinterpretation by the public. Brian et al. (2007) analyse the effect of publishing forecasts through, for example, Fan charts by nine central banks on communicating the central banks monetary policy to the public and in conducting their

monetary policy.

Wieland & Wolters (2013) provide an empirical analysis to show whether, regarding interest rates, policy makers decisions for the Federal Reserve bank and the European Central Bank adjust to forecasts directly. They evaluate policy rules that were made basing on forecasts and find evidence of the interest rate having been set or changed in a way that is in alignment with a response to forecast changes. A very recent application that demonstrates the importance of forecasting for policy is in the prediction of the effect of the Covid 19 crisis on for example GDP by many governments.

3.3 Model Specification, Likelihood and Volatility Estimates

The model that we analyze is the same as in Leon-Gonzalez (2019) and assumes that the distribution of the one dimensional y_t conditional on an observed predetermined vector of regressors x_t can be described as follows:

$$y_t = \mu + x_t\beta + e_t, \qquad e_t|k_t \sim N\left(0, \frac{1}{k_t B^2}\right) \tag{3.3.1}$$

where β is a conformable vector of coefficients, μ and B^2 are scalar parameters and e_t is i.i.d. stationary and independent of x_t. The predetermined regressors x_t are assumed to be either stationary or trend-stationary. The state variable k_t follows an autoregressive Gamma process (Gouriéroux & Jasiak, 2006) which can be described by writing $k_t = z_t' z_t$, where z_t is a $n \times 1$ vector that has the following Gaussian AR(1) representation:

$$z_t = \rho z_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \theta^2 I_n) \tag{3.3.2}$$

where ρ is a scalar that controls the persistence of the volatility, with $|\rho| < 1$ and ε_t is i.i.d. and stationary. The stationary and initial distribution of the time varying inverse volatility k_t is a gamma with n degrees of freedom, such that $k_1 \backsim \text{Gamma}\big(n/2, \frac{2\theta^2}{1-\rho^2}\big)$. Therefore we have that $E(\frac{1}{k_t B^2}) = E(Var(e_t|k_t)) = \frac{1}{B^2}\frac{1-\rho^2}{n-2}$, provided that $n > 2$, where as a normalization we assume $\theta^2 = 1$ because we have B^2 in (3.3.1). For $0 < n \leq 2$ the model

is well-defined but the volatility does not have a finite mean. The conditional distribution of $k_t|k_{t-1}$ is a non central chi-squared times a parameter constant that can be written as a mixture of gammas. The noncentral chi squared is well defined for non-integer values of n, so we will treat the unknown parameter n as a continuous parameter.

Then, given the properties of a gamma, the conditional mean of the inverse volatility k_t given previous history of k_t is a weighted average of the unconditional mean of k_t and its previous value k_{t-1}.

$$E(k_t|k_{t-1}) = \rho^2 k_{t-1} + (1-\rho^2)E(k_t)$$

where $\rho^2 k_{t-1}$ represents the non centrality parameter. k_t is correlated with its previous value and this generates the persistence in the squared residuals, a characteristic feature of time-varying variance models.

The inverse gamma specification implies a student-t distribution with n degrees of freedom for y_t thus enabling us to model heavy tailed distributions. In contrast, the gamma SV model (Creal, 2017) implies a variance-gamma distribution, which has thin tails (Madan & Seneta, 1990). The local scale model of Shephard (1994) is non-stationary, unlike ours which is stationary. In addition, the local scale model requires a restriction on the initial distribution for conjugacy (i.e. $\nu = 2\alpha_1$).

Integrating out analytically the volatilities in our model not only allows us to get an analytical expression for the likelihood, but also to see the similarity of our model to GARCH models. In particular we can see that the variance at each point in time given previous data is a (nonlinear) function of previous residuals. Using the filtering distributions in subsection 3.3.2, we obtain the following:

- $y_1|k_1 \sim N(\mu + x_1\beta, (B^2k_1)^{-1})$, where k_1 is a gamma. Therefore the first observation is a student-t with n degrees of freedom.

- Similarly for the second observation $y_2|y_1, k_2 \sim N(\mu + x_2\beta, (B^2k_2)^{-1})$, where $k_2|y_1$ is a mixture of gammas. $E(k_2|y_1)$ is a nonlinear function of past residuals.

- For any t, $y_t|y_{t-1}, ..., y_1, k_t \sim N(\mu + x_t\beta, (B^2k_t)^{-1})$, where $k_t|y_{t-1}, ..., y_1$ is a mixture of gammas, whose expected value is a nonlinear function of all past residuals.

Thus, integrating out the volatilities gives a structure similar to GARCH models, but with a different functional form and distribution.

3.3.1 The Likelihood

The following proposition, whose proof is in Appendix A.1.2, gives the likelihood for the model described in equations (3.3.1)-(3.3.2).

Proposition 3.3.1. *Let $e_t = y_t - \mu - x_t\beta$ for $t = 1, ..., T$. The likelihood for the first observation is:*

$$L(y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}2^{\frac{1}{2}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left|B^2e_1^2 + V_1^{-1}\right|^{-\frac{n+1}{2}}V_1^{-\frac{n}{2}}$$

for the second is:

$$L(y_2|y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{(B^2e_2^2+1)^{-\frac{n+1}{2}}}{(1-\delta_2)^{-\frac{n+1}{2}}}\hat{C}_2$$

for the third is:

$$L(y_3|y_2, y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\frac{\Gamma\left(\frac{n+1+2h_2}{2}\right)}{(B^2e_3^2+1)^{\frac{n+1}{2}}}(2S_3)^{\frac{n+1+2h_2}{2}}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_3$$

for the fourth is:

$$L(y_4|y_3, y_2, y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_4}\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\frac{\Gamma\left(\frac{n+1+2h_3}{2}\right)}{(B^2e_4^2+1)^{\frac{n+1}{2}}}(2S_4)^{\frac{n+1+2h_3}{2}}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_4$$

and for any $t \geq 3$ is

$$L(y_t|y_{1:t-1}) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_t}\sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}\frac{\Gamma\left(\frac{n+1+2h_{t-1}}{2}\right)}{(B^2e_t^2+1)^{\frac{n+1}{2}}}(2S_t)^{\frac{n+1+2h_{t-1}}{2}}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_t$$

where:

$$V_1 = (1-\rho^2)^{-1}$$

$$\widetilde{V}_2^{-1} = V_1^{-1} + B^2 e_1^2$$

$$\delta_2 = \rho^2(\widetilde{V}_2^{-1} + \rho^2)^{-1}$$

$$Z_2 = (1 + B^2 e_2^2)^{-1}\delta_2$$

$$\widetilde{C}_{2,h_2} = \frac{[(n+1)/2]_{h_2}}{[n/2]_{h_2}}\left(\frac{1}{2}\rho^2(\widetilde{V}_2^{-1} + \rho^2)^{-1}\right)^{h_2}\frac{1}{h_2!}$$

$$\widetilde{C}_{3,h_3} = \sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+1+2h_2}{2}\right)\frac{[(n+1)/2+h_2]_{h_3}}{[n/2]_{h_3}}\left(\frac{1}{2}\rho^2 S_3\right)^{h_3}\frac{1}{h_3!}(2S_3)^{\frac{n+1+2h_2}{2}}$$

$$c_3 = {}_2F_1\left(\frac{n+1}{2}, \frac{n+1}{2}; \frac{n}{2}; \delta_3\right)\Gamma\left(\frac{n+1}{2}\right)(1-\rho^2 S_3)^{-\frac{n+1}{2}}(2S_3)^{\frac{n+1}{2}}$$

$$\hat{C}_t = {}_2F_1\left(\frac{n+1+2h_{t-1}}{2}, \frac{n+1}{2}; \frac{n}{2}; Z_t\right) \text{ for } t \geq 2 \text{ and where } h_1 = 0$$

for $T \geq t \geq 3$

$$S_t = (1 + B^2 e_{t-1}^2 + \rho^2)^{-1}$$

$$\widetilde{V}_t^{-1} = 1 + B^2 e_{t-1}^2$$

$$Z_t = (B^2 e_t^2 + 1)^{-1} S_t \rho^2$$

$$\delta_t = \left((1-\rho^2 S_t)^{-1} S_t \rho^2 (\widetilde{V}_{t-1}^{-1} + \rho^2)^{-1}\right)$$

and for $T+1 \geq t \geq 4$

$$c_t = \sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}(1-\rho^2 S_t)^{-\frac{n+1+2h_{t-1}}{2}}\Gamma\left(\frac{n+1+2h_{t-1}}{2}\right)(2S_t)^{\frac{n+1+2h_{t-1}}{2}}$$

$$\widetilde{C}_{t-1,h_{t-1}} =$$

$$\sum_{h_{t-2}=0}^{\infty}\widetilde{C}_{t-2,h_{t-2}}\Gamma\left(\frac{n+1+2h_{t-2}}{2}\right)\frac{[(n+1)/2+h_{t-2}]_{h_{t-1}}}{[n/2]_{h_{t-1}}}\left(\frac{1}{2}\rho^2 S_{t-1}\right)^{h_{t-1}}\frac{(2S_{t-1})^{\frac{n+1+2h_{t-2}}{2}}}{h_{t-1}!}$$

and $S_{T+1} = (1 + B^2 e_T^2)^{-1}$

The rising factorial is denoted as $[x]_h$ and ${}_2F_1$ denotes a hypergeometric function (e.g. Muirhead (2005, p. 20)). There are a number of transformations to the ${}_2F_1$ hypergeometric functions above to accelerate their convergence. Abramowitz et al. (1988, p.559) defines several transformations such as the Euler transformation:

$$ {}_2F_1(a,b;c;z) = (1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c;z) $$

or a linear combination approach:

$$ \begin{aligned} {}_2F_1(a,b;c;z) = {} & \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\,{}_2F_1(a,b;a+b-c+1;1-z) \\ & + (1-z)^{c-a-b}\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}\,{}_2F_1(c-a,c-b;c-a-b+1;1-z) \\ & \text{for } (|arg(1-z)| < \pi) \end{aligned} $$

The expression for $\hat{C}_t$ above transformed using the Euler transformation becomes:

$$ \hat{C}_t = (1-Z_t)^{-\frac{n+2+2h_{t-1}}{2}}\,{}_2F_1\left(-\frac{1+2h_{t-1}}{2}, -\frac{1}{2}; \frac{n}{2}; Z_t\right) \text{ for t} \geq 2 \text{ and where } h_1 = 0 $$

In our coding we used the Euler acceleration only for $\hat{C}_2$ and c_3, because for larger values of t the acceleration did not converge when h was large. Regarding the linear combination approach, although we did not implement it in our code for the R package, the acceleration converges. We accelerated the calculations by implementing parallel computing in the code. This is possible because many of the coefficients in the series are the same for every t, therefore they only need to be computed once, which can be done in parallel. We also calculate all the $\hat{C}_t$ in parallel. As shown in Section 3.4, this drastically reduces computation time. The derivatives of the log-likelihood can be obtained as a byproduct of the likelihood calculation.

After integrating out the volatilities, this likelihood can be calculated recursively starting with y_1, which is the first observation, to y_T. This likelihood is easy to compute and it always converges since $|Z_t| < 1$ for all values of t. We truncate the number of terms to calculate the hypergeometric functions to around 350 to ensure convergence, and the sums are truncated at about $h = 350$. These truncation values seemed to be sufficient as

explained in Table 1 in our application using inflation data.

3.3.2 Joint Smoothing and Filtering Distributions

In this subsection, we provide the analytical expressions for both the joint smoothing and filtering distributions for the volatilities. Propositions 3.3.2, 3.3.3 and 3.3.4, proved in the Appendix, provide the smoothing distributions in alternative forms. Proposition 3.3.2 and 3.3.3 give the conditional distributions $\pi(k_t|k_{(t+1):T}, y_{1:T})$, and $\pi(k_t|k_{1:(t-1)}, y_{1:T})$, respectively, while Proposition 3.3.4 gives the marginals $\pi(k_t|y_{1:T})$. The filtering distributions are stated after Proposition 3.3.4.

Proposition 3.3.2. *The joint posterior distribution $\pi(k_{1:T}|y_{1:T})$ can be obtained from the following conditional densities each of which is a mixture of gammas:*

$$\pi(k_t|k_{(t+1):T}, y_{1:T}) \propto |k_t|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_{t+1}^{-1}k_t\right) \sum_{h=0}^{\infty} \left(C_{t,h}|k_t|^h\right),\ t = 1, ..., T$$

where

$$C_{1,h} = \frac{1}{h!}\frac{1}{[n/2]_h}\left(\frac{1}{4}\rho^2 k_2\right)^h$$
$$S_2 = (1 + B^2 e_1^2)^{-1}$$
$$S_{T+1} = (1 + B^2 e_T^2)^{-1}$$

for $3 \leq t \leq T$

$$S_t = (1 + B^2 e_{t-1}^2 + \rho^2)^{-1}$$

and for $2 \leq t < T$:

$$C_{t,h} = \sum_{h_t=0}^{h} \widetilde{C}_{t,h-h_t} \frac{1}{[n/2]_{h_t}} \left(\frac{1}{4}\rho^2\right)^{h_t} \frac{k_{t+1}^{h_t}}{h_t!}$$

while for $t = T$, $C_{t,h} = \widetilde{C}_{t,h}$, and where $\widetilde{C}_{t,h}$ has been defined in Proposition 3.3.1.

Proposition 3.3.3. *The density $\pi(k_t|k_{1:(t-1)}, y_{1:T})$ is a mixture of gamma distributions and its kernel is proportional to:*

$$\pi(k_{\text{T-s}}|k_{1:(\text{T-s-1})}, y_{1:T}) \propto |k_{\text{T-s}}|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S^{-1}_{\text{T-s}+1}k_{\text{T-s}}\right)\left(\sum_{h=0}^{\infty} a_{\text{T-s},h}k^h_{\text{T-s}}\right) \qquad s = 0, \ldots, T-1$$

where

$$a_{\text{T-s},h} = \sum_{h_{\text{T-s}}=0}^{h} \widetilde{a}_{\text{T-s},h-h_{\text{T-s}}} \frac{1}{(h_{\text{T-s}})!}\frac{1}{[n/2]_{h_{\text{T-s}}}}\left(\frac{1}{4}\rho^2\right)^{h_{\text{T-s}}} k^{h_{\text{T-s}}}_{\text{T-s}-1}, \qquad s = 1, \ldots, T-2$$

and

$$\begin{aligned}\widetilde{a}_{\text{T-s},h_{\text{T-s}+1}} = &\sum_{h_{\text{T-s}+2}=0}^{\infty} \widetilde{a}_{\text{T-s}+1,h_{\text{T-s}+2}} \Gamma\left(\frac{n+1}{2} + h_{\text{T-s}+2}\right)\frac{[(n+1)/2 + h_{\text{T-s}+2}]_{h_{\text{T-s}+1}}}{[n/2]_{h_{\text{T-s}+1}}} \\ &\times \frac{\left(\frac{1}{2}\rho^2 S_{\text{T-s}+2}\right)^{h_{\text{T-s}+1}}}{(h_{\text{T-s}+1})!}(2S_{\text{T-s}+2})^{\frac{n+1+2h_{\text{T-s}+2}}{2}} \qquad s = 2, \ldots, T-1\end{aligned}$$

with,

$$a_{T,h} = \frac{1}{h!}\frac{1}{[n/2]_h}\left(\frac{1}{4}\rho^2 k_{T-1}\right)^h$$

$$\widetilde{a}_{T-1,h_T} = \frac{[(n+1)/2]_{h_T}}{[n/2]_{h_T}}\frac{\left(\frac{1}{2}\rho^2 S_{T+1}\right)^{h_T}}{(h_T)!}$$

For the case when $s = T-1$, we have $a_{\text{T-s},h} = a_{1,h} = \widetilde{a}_{1,h}$.

We can integrate $\pi(k_{1:T})\pi(y_{1:T}|k_{1:T})$ with respect to $k_{1:(t-1)}$ and with respect to $k_{(t+1):T}$ to obtain the following proposition which gives the marginal density $\pi(k_t|y_{1:T})$ for $t = 2, \ldots, T-1$. Note that for $t = T$ or $t = 1$ the marginal densities are given by Propositions 3.3.2 and 3.3.3, respectively.

Proposition 3.3.4. *The density of $\pi(k_t|y_{1:T})$ is that of a mixture of gammas and its kernel is given by:*

$$\pi(k_t|y_{1:T}) \propto |k_t|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S^{-1}_{t+1}k_t\right)\sum_{h=0}^{\infty}\widetilde{D}_{t,h}|k_t|^h$$

for $t = 2, \ldots, T-1$, where for $2 \le t < T-1$:

$$\widetilde{D}_{t,h} = \sum_{h_t=0}^{h}\sum_{h_{t+1}=0}^{\infty}\widetilde{C}_{t,h-h_t}\frac{1}{[n/2]_{h_t}}\left(\frac{1}{4}\rho^2\right)^{h_t}\frac{1}{h_t!}\frac{\Gamma\left(\frac{n+1}{2} + h_t + h_{t+1}\right)}{(S^{-1}_{t+2}/2)^{\frac{n+1}{2}+h_t+h_{t+1}}}\widetilde{a}_{t+1,h_{t+1}}$$

and for $t = T-1$*:*

$$\widetilde{D}_{T-1,h} = \sum_{h_{T-1}=0}^{h} \widetilde{C}_{T-1,h-h_{T-1}} \frac{1}{[n/2]_{h_{T-1}}} \left(\frac{1}{4}\rho^2\right)^{h_{T-1}} \frac{1}{(h_{T-1})!} \frac{\Gamma\left(\frac{n+1}{2} + h_{T-1}\right)}{(S_{T+1}^{-1}/2)^{\frac{n+1}{2}+h_{T-1}}}$$

where $\widetilde{a}_{t+1,h}$ *was defined in Proposition 3.3.3 and* $\widetilde{C}_{t,h_t}$ *was defined in Proposition 3.3.1.*

Regarding the filtering distributions, they were obtained in the proof of Proposition 3.3.1. They are a mixture of gammas and the kernel is given by:

$$\pi(k_t|y_{t-1}, y_{t-2}, ..., y_1) \propto |k_t|^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_t\right) \sum_{h=0}^{\infty} \left(\widetilde{C}_{t,h}|k_t|^h\right),\ t = 1, ..., T$$

where the recursive constants are defined in Proposition 3.3.1.

3.4 Empirical Applications

3.4.1 Macroeconomic Data

To illustrate the efficiency and usefulness of our proposed novel addition to the SV literature, we provide macroeconomic applications using inflation data for the UK, Japan, US and Brazil. The data series were all sourced from the Federal Reserve Bank of St Louis Fred database as the Consumer Price Index (CPI) data and inflation was constructed using the following formula:

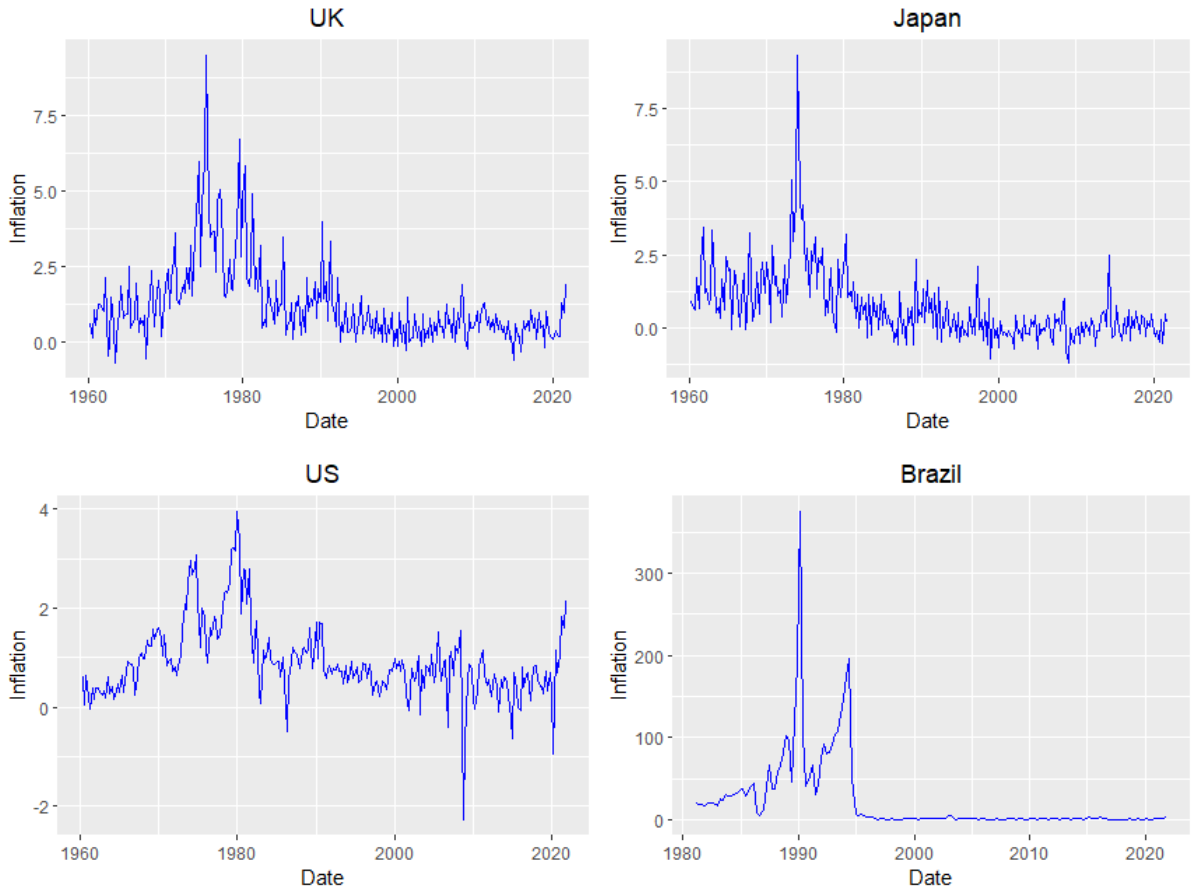
$$\text{Inflation} = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} \times 100$$

The number of observations for each series were determined by availability of data. UK data thus covers the period 1960Q2 to 2022Q1 and Japan data is obtained for the period 1960Q4 to 2022Q1. The US inflation data covers the period 1960Q1 to 2021Q4. Due to unavailability of data for earlier years for Brazil we have observations for the period 1981Q1 to 2021Q4. y_t is the level of inflation and x_t contains a constant and 4 lags of y_t. Therefore, for each series we have 244, 242, 244 and 160 observations, respectively, after

constructing the lags.

Figure 3.1 illustrates the quarterly inflation series for the four countries in levels. The trend for the evolution of inflation for the US, UK and Japan in the early 1970's and 1980's have slight similarities. However, in later years across all series, inflation evolves differently.

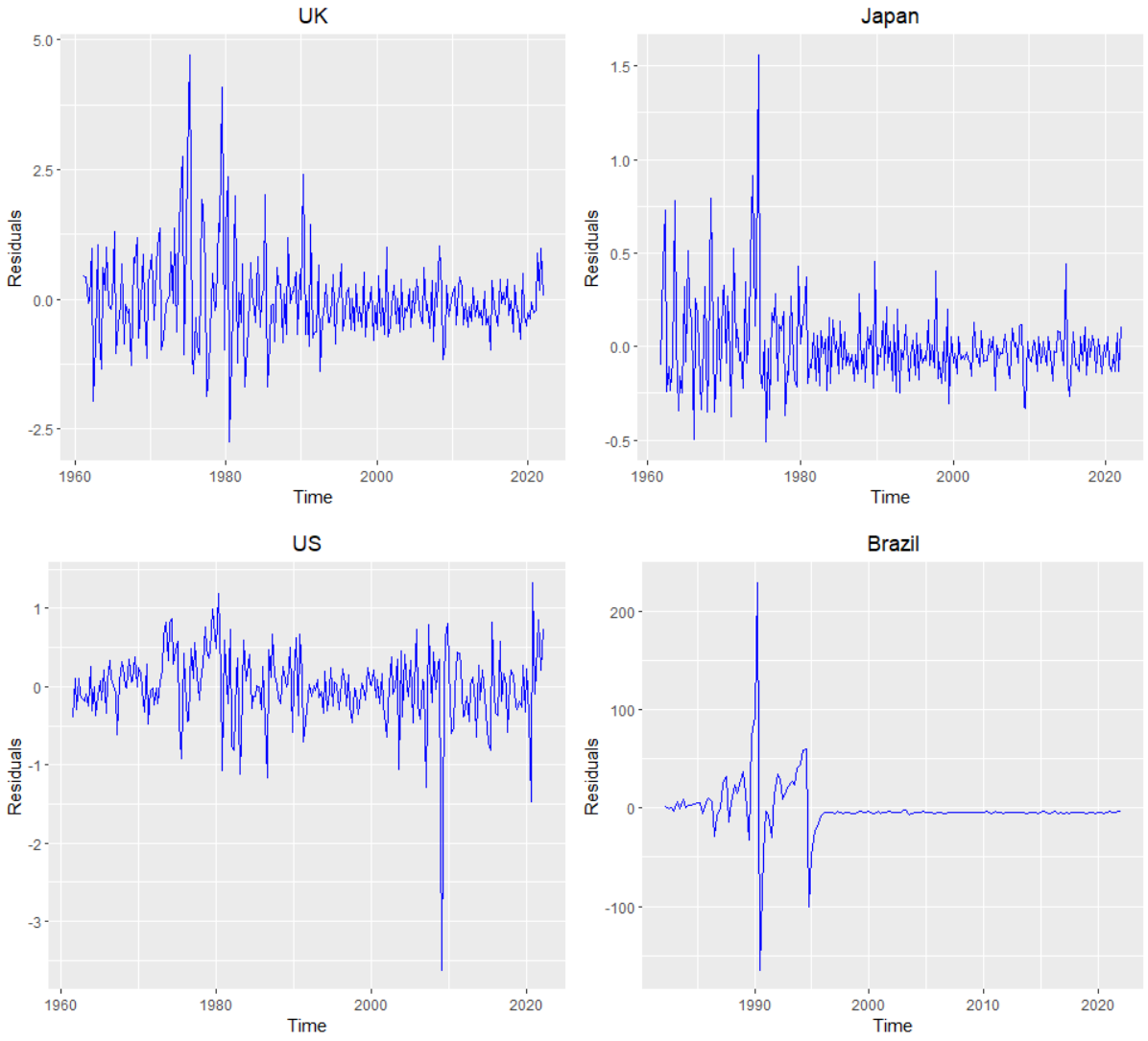
Figure 3.1: Inflation Rates



The x-axis plots the dates that correspond to the end of each year for the quarterly observations. The y-axis plots the Inflation Rates

Figure 3.2 shows the Ordinary Least Squares (OLS) residuals for the four countries over the sample period, after regressing the level of inflation on its 4 lags and an intercept. The spikes in volatility observed for Brazil inflation show that the series accumulates periods of consistent high volatility continuously. Overall for all countries, volatility patterns exhibit some extreme values suggesting that models that assume heavier tailed distributions might fit better and improve forecasting.

Figure 3.2: Residuals Plots



The x-axis plots the time period. The y-axis plots the OLS Residuals

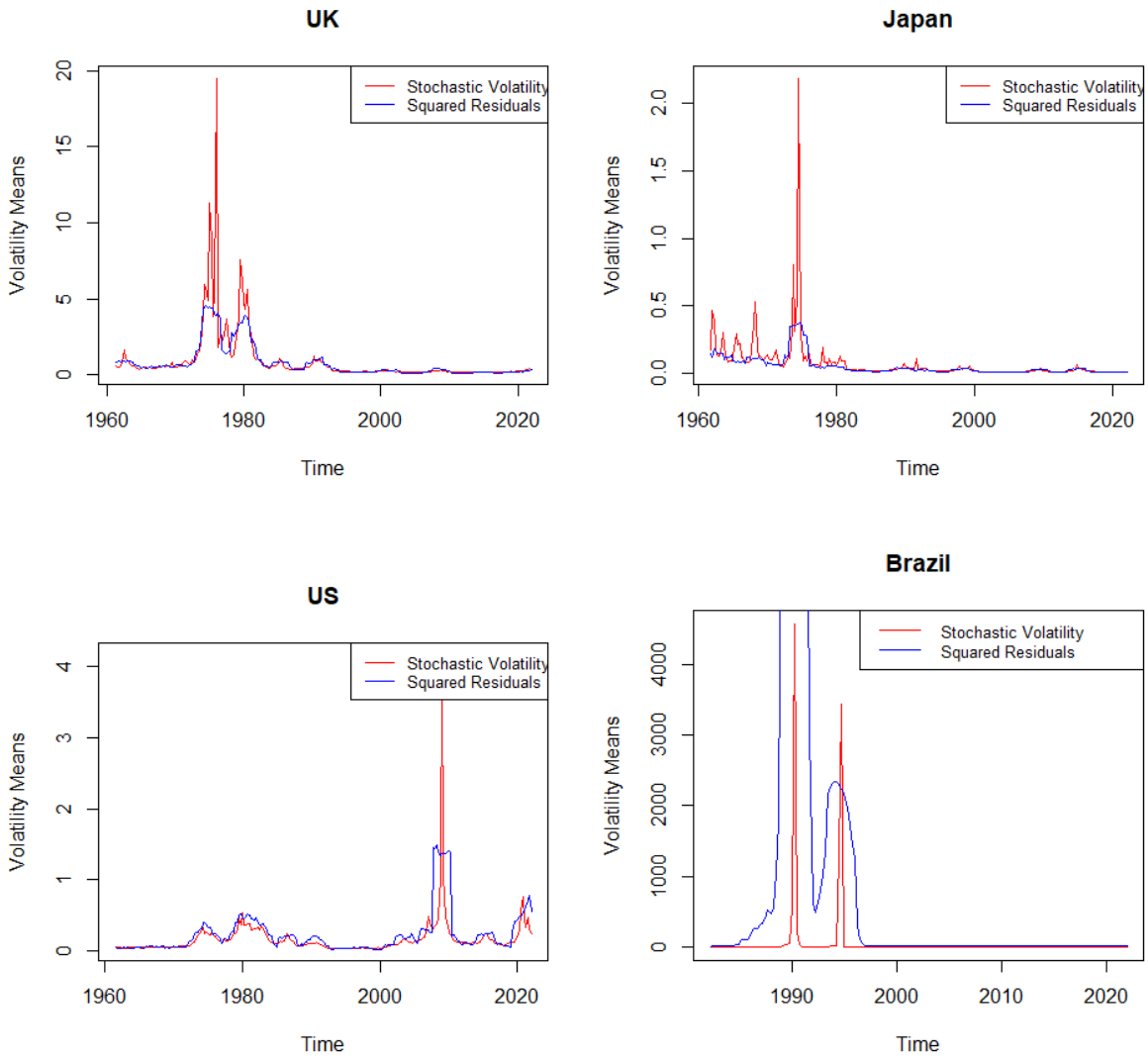
In the maximization algorithm, the initial values for the slope coefficients are equal to the OLS estimates, and for the rest of the parameters we choose values such that the mean volatility implied by the model equals that of the data. We truncate the calculation of hypergeometric functions at 350 terms and we truncate h_t in the likelihood at $h_t = 350$ to ensure convergence.

Smoothed Estimates of the Volatilities

Using the smoothing distributions we are able to obtain an estimate of the variance of e_t at each point of time given all available data: $E(var(e_t|k_t)) = E(var(y_t|x_t, k_t))$, where

the expectation is with respect to the smoothing distribution of k_t (i.e. $\pi(k_t|y_{1:T})$). This is in contrast to the commonly used GARCH MLE estimates, which can only provide the filtered estimates of the variance: $var(e_t|y_{1:(t-1)})$. Figure 3.3 compares the MLE smoothed estimates of the variance at each point in time for each country, to the moving average of the squared OLS residuals obtained from 5 contiguous squared residuals.

Figure 3.3: Smoothed Estimates of the Volatilities



The red lines show the smoothed estimates of the volatilities compared to the moving average of OLS squared residuals displayed in blue

The periods with high residuals coincide with periods of high estimated stochastic volatility each point in time for all the four countries. In particular for the US and UK the

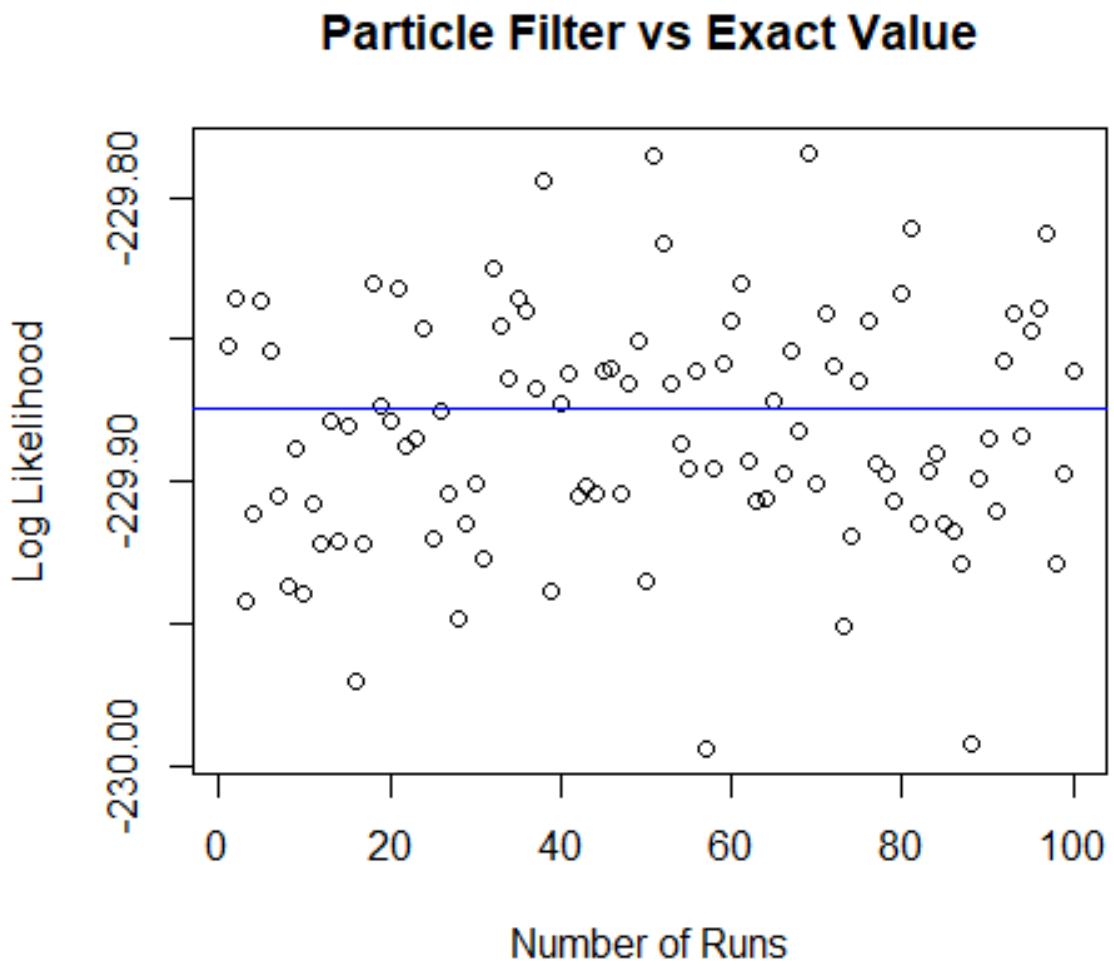
estimates reflect the expectations for high volatility trends observed during periods such as the Great recession and smaller peaks in volatility representing the covid recessions.

Accuracy Check

We compare our novel algorithm to the Particle Filter to check the accuracy of our computations. Particle filters are commonly used in practice for calculating the likelihood function. Literature has it that they provide an unbiased estimate of the likelihood (see e.g Moral (2004), proposition 7.4.1). We use the UK inflation data for this exercise. Parameter values for both algorithms are set at the maximum likelihood estimates. To evaluate each value of the likelihood we use the average of 110 independent replications of the particle filter proposed in Chan et al. (2020). We set the number of particles to twice the sample size T, that is each particle filter has $T * 2$ particles. We obtained 100 values for the log-likelihood using this method and plot them in Figure 3.4 together with the value provided by our algorithm.

The exact log likelihood estimate for the UK inflation data is -229.87. The figure shows that the particle filter value for the log-likelihood goes above and below our exact value. Therefore our solution seems accurate.

Figure 3.4: Particle Filter Estimates UK Inflation



The horizontal blue line represents the exact value obtained using our novel algorithm. Small circles show the 100 log-likelihood estimates, each of which was obtained by averaging 110 runs of the particle filter

Computational Efficiency

In order to calculate the likelihood, we need to truncate the number of terms that are added for the hypergeometric functions ($niter$), and also we need to truncate h. For simplicity we use the same truncation points for both. Table 3.1 shows the values of the log likelihood obtained for several truncation values, using the MLE estimates for the parameter values and the four datasets. The value of the log-likelihood remains stable at

truncation points of 150 (Japan), 200 (US), 300 (UK) and 350 (Brazil)[6].

Using a truncation point of 350, the computation time for one evaluation of the likelihood in seconds for the UK inflation dataset ($T = 244$) is 0.24, 0.39, 0.72 and 2.60 when using 18, 8, 4, or just one computing thread, respectively. For the UK exchange rate dataset ($T = 999$) that we use in Section 4.2 a truncation point of 350 was also adequate, and the computation times for the same increase to 0.82, 1.42, 2.72, 10.07, respectively. The coding was done in C++, linked to the R software and executed in a Ryzen threadripper 3970x processor.

Table 3.1: Likelihood at different truncation parameter values

	UK	**Japan**	**US**	**Brazil**
$niter = h = 100$	-234.59	102.58	-124.61	-392.51
$niter = h = 150$	-230.48	102.67	-124.58	-387.29
$niter = h = 200$	-229.91	102.67	-124.57	-385.91
$niter = h = 300$	-229.87	102.67	-124.57	-385.63
$niter = h = 350$	-229.87	102.67	-124.57	-385.62
$niter = h = 400$	-229.87	102.67	-124.57	-385.62

Parameter Estimates and Model Comparison

Maximum likelihood parameter estimates are reported in Table 3.2 for our model using quarterly inflation data for the UK, Japan, US and Brazil and their standard errors in parenthesis. β_0 is the coefficient of the intercept while $\beta_{1:4}$ are the coefficients of the lags. Throughout the maximum likelihood estimation, we imposed the constraint $0 < \rho < 1$ on the persistence of volatility.

[6]The determining factors for the truncation point, are the sample size. That is, the larger the sample size the larger the truncation point, albeit not linearly. In addition, data with fat tails e.g. that of Brazil requires a larger truncation point. Lastly, when n is close to 0 and rho is close to 1, the truncation point required would be larger

Table 3.2: Inverse Gamma SV Model Maximum Likelihood Estimates

Parameter	**UK**	**Japan**	**US**	**Brazil**
B^2	0.0653	2.2868	0.2845	0.0127
	(0.0354)	(1.2679)	(0.1670)	(0.0064)
ρ	0.9849	0.9734	0.9577	0.9964
	(0.0091)	(0.0159)	(0.0252)	(0.0048)
n	2.2527	2.0529	3.2136	0.7010
	(0.6534)	(0.4724)	(0.8377)	(0.1374)
β_0	0.1148	0.0053	0.1053	-0.1030
	(0.0492)	(0.0078)	(0.0418)	(0.0810)
β_1	0.1256	0.0222	0.5772	1.0604
	(0.0529)	(0.0557)	(0.0701)	(0.0607)
β_2	0.1627	0.2592	0.0500	-0.4053
	(0.0479)	(0.0537)	(0.0731)	(0.0499)
β_3	-0.1005	0.0247	0.3304	0.4889
	(0.0483)	(0.0517)	(0.0719)	(0.0924)
β_4	0.6140	0.4291	-0.0747	-0.0652
	(0.0485)	(0.0530)	(0.0638)	(0.0315)

The coefficients of the lags are mostly significant, and the estimates of ρ indicate high persistence of the volatility in all countries. In all cases except Brazil, the estimated values of n are bigger than 2, implying a finite value for the expected value of volatility. For Brazil we have $n = 0.7$, implying that y_t has very fat tails, similar to those of a Cauchy distribution.

We compare the empirical performance of the following 7 models:

M_1: Homoscedastic

M_2: Local scale model (Shephard, (1994))

M_3: Univariate GARCH(1,1) with normal errors

M_4: Univariate GARCH(1,1) with student t errors

M_5: Log Normal stochastic volatility (e.g. Kim et al. (1998))

M_6: Gamma stochastic volatility

M_7: Inverse Gamma stochastic volatility

Except M_5 all models are estimated by MLE. The model M_5 is estimated using Bayesian methods with the R package *stochvol* (Kastner (2016)), using the default non-informative priors implemented in the package. For this model the value of the log-likelihood at the posterior mean of parameters is evaluated by averaging 50 independent replications of a bootstrap particle filter, with each particle filter having a number of particles equal to 60 times the sample size. Thus, the model is estimated at the posterior mean of the parameters and not the posterior mode. It is possible that estimating the likelihood at the posterior mode of the parameters may increase the value of the likelihood. The numerical standard error of the log-likelihood estimate was smaller than 0.02 in all cases. Both the Gaussian and Student t GARCH are specified as $GARCH(1,1)$, thus they have 8 parameters and 9 parameters respectively given that we have 4 lags and an intercept. The stochastic volatility models have 8 parameters except for the gamma SV model which has an additional parameter for the skewness of volatility.

Table 3.3 reports the log likelihood values at the maximum likelihood estimates and Table 3.4 reports the values of the Bayesian Information Criterion (BIC, Schwarz 1978). As expected the homoscedastic model is the worst of all models for all countries. In terms of the log-likelihood the inverse gamma model is the best for the US, and the gamma SV model is the best for the UK and Japan. For Brazil the GARCH(1,1) with student-t errors has the best value of the log-likelihood, but when penalizing for the number of parameters using the BIC (the smaller the better) the inverse gamma SV model is the best. In summary, using the BIC the gamma SV model is the best for the UK and Japan, and the inverse gamma SV model is the best for the US and Brazil. In the case of the UK and Japan the asymmetry parameter of the Gamma SV model was estimated to be large, which might be the reason for the better performance of this model. In the case of Brazil and the US the residuals appear to have more abrupt changes, which might be the reason for the better performance of the inverse Gamma SV model.

Table 3.3: Inflation Rates Model Comparisons: Log Likelihood

Model	**UK**	**Japan**	**US**	**Brazil**
M_1	-306.74	18.39	-165.42	-763.33
M_2	-230.04	100.17	-129.90	-395.30
M_3	-233.01	90.87	-147.72	-387.76
M_4	-227.74	107.06	-133.34	-383.97
M_5	-229.08	101.96	-126.74	-389.63
M_6	-220.88	112.09	-129.33	-475.07
M_7	-229.87	102.67	-124.57	-385.62

Table 3.4: Inflation Rates Model Comparisons: BIC

Model	**Parameters**	**UK**	**Japan**	**US**	**Brazil**
T		244	242	244	160
M_1	6	646.46	-3.85	363.83	1557.12
M_2	8	504.05	-156.42	303.77	831.21
M_3	8	509.99	-137.83	339.41	816.11
M_4	9	504.95	-164.73	316.15	813.61
M_5	8	502.13	-160.00	297.47	819.85
M_6	9	491.24	-174.79	308.14	995.81
M_7	8	503.72	-161.43	293.12	811.84

3.4.2 Exchange Rates Data Application

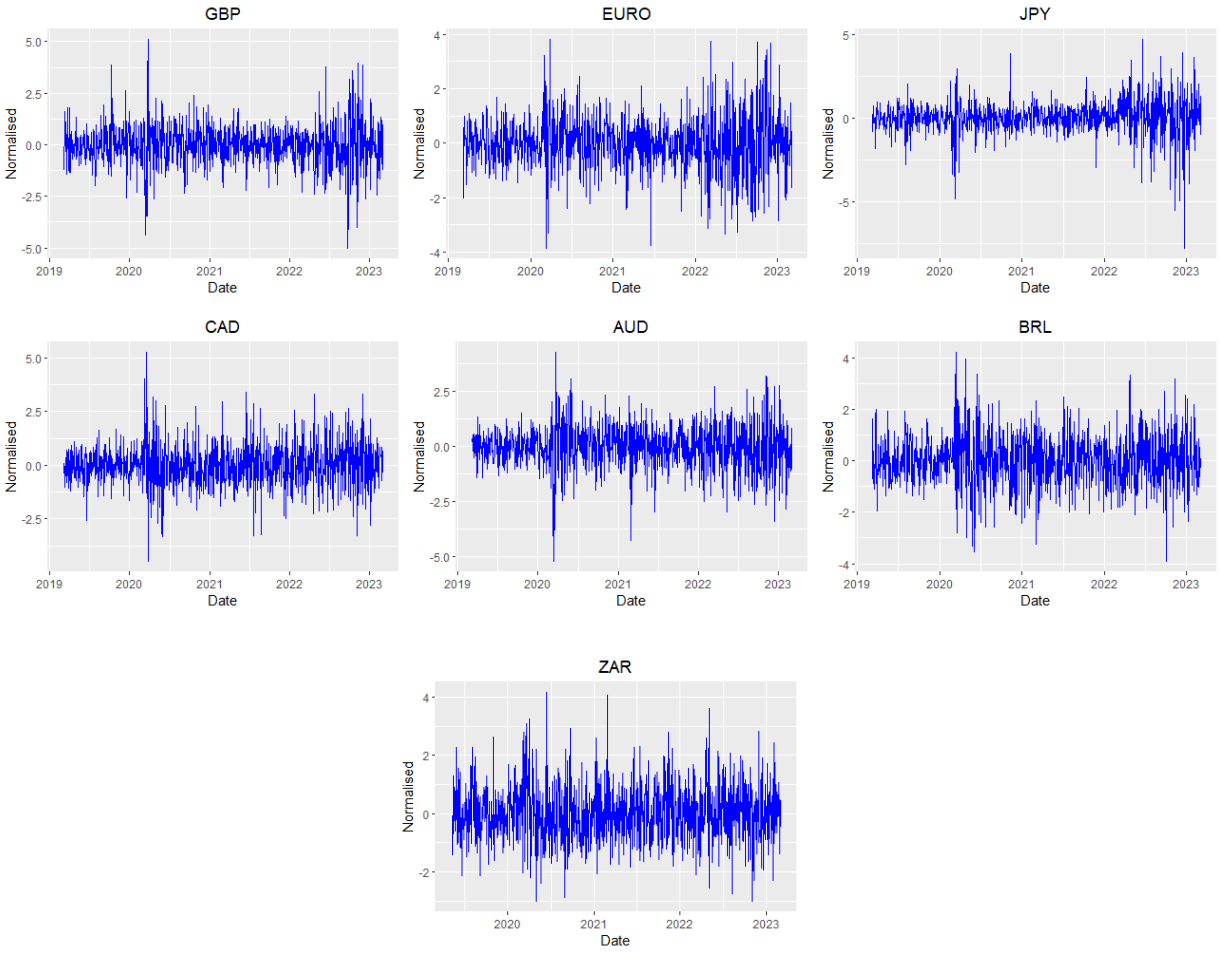
We use 1000 daily exchange rate observations for 7 currencies (GBP, EUR, JPY, CND, AUD, BRL, ZAR) to the USD. The data for the first 6 currencies were obtained from the Board of Governors of the Federal Reserve and covers the period beginning 5 March 2019 and ending 3 March 2023. ZAR was obtained from the South African Reserve Bank for the period 7 May 2019 to 3 March 2023. In this analysis y_t is the first differences of the log exchange rate. All models include an intercept but we include no regressors (i.e.

x_t is empty).

Figure 3.5 shows the normalised exchange rates for the 7 countries. We calculate the percentage of times that the absolute value of the normalised exchange rate goes beyond 1.96 standard deviations. The JPY, BRL, GBP, CAD, EUR, and AUD have thicker tails than a normal distribution with 5.8%, 5.7%, 5.9%, 5.2%, 6.5% and 6.1% proportions respectively. The ZAR has slightly thinner tails to the normal with 4.8% of the proportion going beyond 1.96 standard deviations.

In addition, we obtain the proportion where the absolute value of the normalised exchange rate goes beyond 3.0902 standard deviations, which is 0.2% for a normal distribution. The ZAR has the lowest proportion, with 0.4%, but still larger than the normal. The JPY, BRL, GBP, CAD, EUR, and AUD distribution proportions are 1.8%, 1.0%, 1.6%,1.3%, 1.2%, 0.9%, respectively, all of them much greater than the normal.

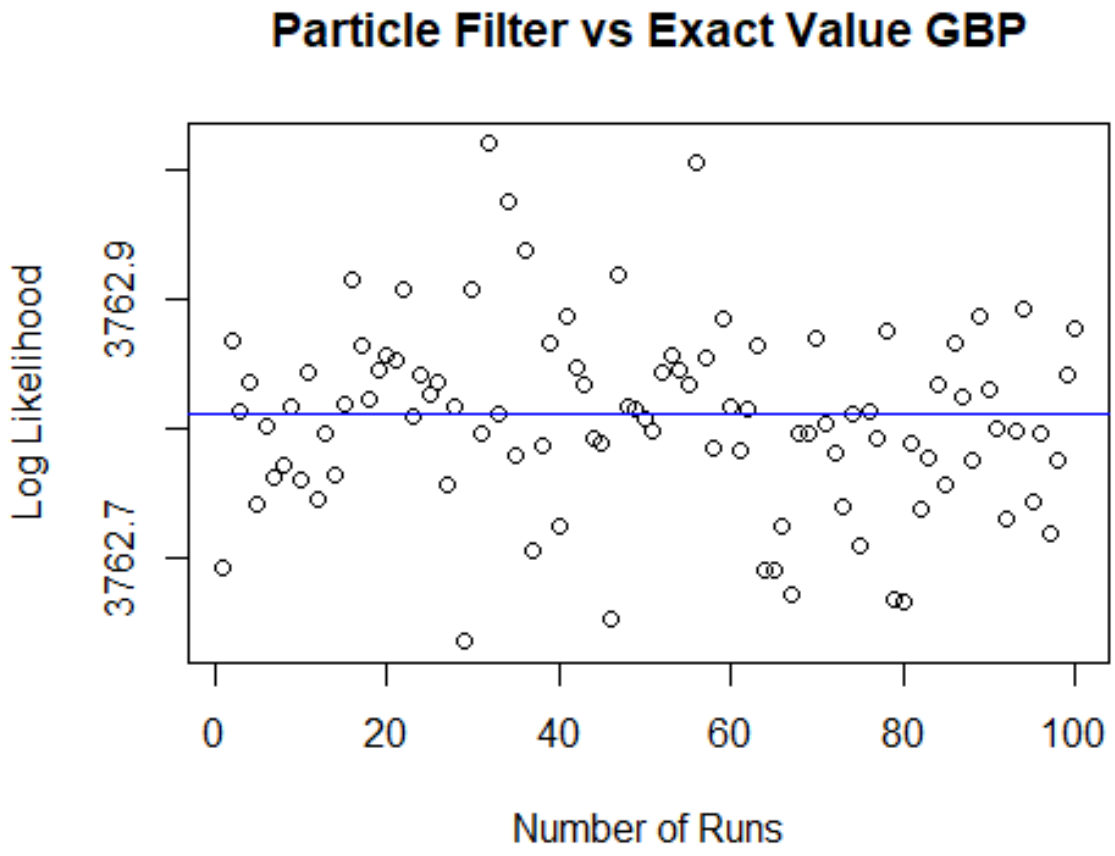
Figure 3.5: Normalised Exchange Rates



y_t was normalised by subtracting its mean and dividing by the standard deviation. The x-axis plots the dates that correspond to the end of each year for the daily observations. The y-axis plots the normalised y_t

When the same size is larger, the approximation error tends to accumulate. To check the accuracy of our novel algorithm over a larger sample size, we do the same exercise as with the UK inflation data, for the exchange rates data which has 999 observations for the Great British Pound. Parameter values for both algorithms are set at the maximum likelihood estimates as before. Figure 3.6 plots the result. The value of the log likelihood using our model is 3762.81. The figure shows that the particle filter value for the log-likelihood goes above and below our exact value as before even with a larger sample size.

Figure 3.6: Particle Filter Estimates GBP



The horizontal blue line represents the exact value obtained using our novel algorithm. Small circles show the 100 log-likelihood estimates, each of which was obtained by averaging 110 runs of the particle filter

Similarly, with a larger sample size, we expect that the truncation point will be larger, the larger the sample size is. To see by how much this increase tends to be we again evaluate the likelihood at different truncation parameter values for the Exchange rates data. We begin with the truncation

Table 3.5: Likelihood at different truncation parameter values Exchange Rates

Model	**GBP**	**EUR**	**JPY**	**CAD**	**AUD**	**BRL**	**ZAR**
$niter = h = 200$	3762.05	4022.94	3973.79	4023.11	3618.08	3152.75	3251.88
$niter = h = 250$	3762.74	4040.93	3973.79	4032.97	3631.00	3163.01	3251.90
$niter = h = 300$	3762.81	4049.34	3973.79	4036.66	3636.62	3166.89	3252.42
$niter = h = 350$	3762.81	4053.12	3973.79	4037.86	3638.98	3168.27	3252.42
$niter = h = 400$	3762.81	4054.69	3973.79	4038.17	3639.87	3168.70	3252.42
$niter = h = 450$	3762.81	4055.26	3973.79	4038.35	3640.14	3168.81	3252.42
$niter = h = 500$	3762.81	4055.44	3973.79	4038.36	3640.21	3168.84	3252.42
$niter = h = 550$	3762.81	4055.48	3973.79	4038.36	3640.23	3168.94	3252.42
$niter = h = 600$	3762.81	4055.50	3973.79	4038.36	3640.23	3168.94	3252.42

Table 3.5 shows the values of the log likelihood obtained for several truncation values, using the MLE estimates for the parameter values and the seven data sets. The value of the log-likelihood remains stable at truncation points of 300 (GBP), 600 (EURO), 200 (JPY), 500 (CAD), 550(AUD), 600 (Brazil) and, 300 (ZAR). The increase in the truncation point is not linearly proportional to the increase in sample size. For example, the JPY data is stable after just 200 points.

Table 3.6 shows the log likelihood values and Table 3.7 the BIC values (the smaller the better) across all the 7 models listed above. The best model for the ZAR, which has the thinnest tails, is the Gamma SV model, both in terms of the likelihood and the BIC. For all the other currencies the GARCH(1,1) with student-t errors has the highest log-likelihood values. However, when taking into account the number of parameters using the BIC, this model is the best only for the EUR and JP. The inverse Gamma SV model is the best for all the other currencies, GBP, CAD, AUD, BRL, with the log normal SV model being equally good for the GBP and BRL.

Table 3.6: Exchange Rates Model Comparisons: Log likelihood

Model	**GBP**	**EUR**	**JPY**	**CAD**	**AUD**	**BRL**	**ZAR**
M_1	3659.66	3962.76	3770.79	3976.17	3551.11	3123.64	3236.27
M_2	3754.46	4053.28	3962.96	4034.74	3637.23	3167.60	3244.68
M_3	3747.26	4044.27	3927.91	4027.51	3632.31	3165.52	3249.98
M_4	3765.21	4059.93	3987.26	4041.50	3641.47	3171.54	3253.80
M_5	3762.50	4055.48	3971.76	4036.88	3638.15	3168.97	3251.93
M_6	3759.35	4053.41	3967.96	4034.91	3633.98	3168.72	3257.63
M_7	3762.81	4055.50	3973.79	4038.36	3640.23	3168.94	3252.42

Table 3.7: Exchange Rates Model Comparisons: BIC

Model	**Parameters**	**GBP**	**EUR**	**JPY**	**CAD**	**AUD**	**BRL**	**ZAR**
T		999	999	999	999	999	999	999
M_1	2	-7305.51	-7911.71	-7527.77	-7938.52	-7088.41	-6233.46	-6458.73
M_2	4	-7481.30	-8078.92	-7898.28	-8041.86	-7246.83	-6307.58	-6461.73
M_3	4	-7466.89	-8060.91	-7828.19	-8027.39	-7236.99	-6303.42	-6472.33
M_4	5	-7495.89	-8085.33	-7939.98	-8048.47	-7248.40	-6308.55	-6473.07
M_5	4	-7497.37	-8083.33	-7915.89	-8046.13	-7248.67	-6310.31	-6476.23
M_6	5	-7484.16	-8072.29	-7901.38	-8035.29	-7233.42	-6302.92	-6480.73
M_7	4	-7497.99	-8083.37	-7919.96	-8049.09	-7252.83	-6310.25	-6477.22

3.4.3 Stochastic Volatility Model with Leverage

Omori et al. (2007) suggest an approach that can handle stochastic volatility models with leverage effects. Regarding the log normal stochastic volatility model, the model proposed by Omori et al. (2007) has a similar structure to that of the baseline approach by Kim et al. (1998) above, where in this case e_t and δ_t are both gaussian as follows:

$$\begin{pmatrix} e_t \\ \delta_t \end{pmatrix} \backsim \mathrm{N}(0, \Sigma)$$

where:

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}$$

The leverage effect is thus captured by the parameter ρ for this class of non linear state space models. Literature is replete with papers that find that this leverage effect is particularly a key property of high frequency financial data. In what follows, we compare our novel inverse gamma stochastic volatility model to the model proposed by Omori et al. (2007).

We use daily returns for the top 4 stocks each on the Tokyo Stock Price Index (Topix) and the Standards and Poors 500 ($S\&P500$), which are ranked by market capitalisation as of November 2023. The 4 stocks with the assigned stock codes or ticker symbols for TOPIX are Toyota ($7203.T$), Keyence Corporation $6861.T$, Sony($6758.T$) and, Nippon Telegraph and Telephone Corporation($9432.T$). The $S\&P500$ stocks are APPLE(AAPL), Microsoft Corporation(MSFT),NVIDIA Corporation(NVDA) and Amazon(AMZN). The stock returns are obtained from Yahoo finance, and consist of 1000 observations for each historical return for the period 29 November 2023 to 17 November 2023. The observations are transformed by taking first differences of the log returns.

The asymmetric stochastic volatility model with leverage is estimated using Bayesian methods with the R package *asv* (Omori (2024)), using the default non-informative priors implemented in the package. For this model, the model parameters are estimated using MCMC methods. Then, the value of the log-likelihood, given the estimated parameters, is evaluated using an auxiliary particle filter provided with the package with 5000 particles. The number of simulations used throughout the application was 150000 with a burn in of 15000. We also add the Gamma stochastic volatility model and the log normal stochastic volatility models above to the comparison. The models have either 4 or 5 parameters given that we have an intercept. Table 3.8 shows the log likelihood values and the BIC values (the smaller the better) for these models.

Table 3.8: Log Likelihood and BIC Comparison against SV model with leverage

	Log Likelihood				BIC			
Model	**IGSV**	**ASV**	**GammaSV**	**LNSV**	**IGSV**	**ASV**	**GammaSV**	**LNSV**
Parameters	**4**	**5**	**5**	**4**	**4**	**5**	**5**	**4**
AAPL	2574,25	2572,55	2566,46	2572,60	−5120,87	−5110,57	−5098,38	−5117,57
MSFT	2611,36	2615,31	2605,87	2612,21	−5195,08	−5196,08	−5177,19	−5196,79
AMZN	2413,43	2413,19	2409,46	2412,35	−4799,23	−4791,85	−4784,38	−4797,07
NVDA	2052,70	2053,87	2051,91	2053,66	−4077,77	−4073,20	−4069,28	−4079,69
6861.T	2564,39	2565,19	2565,00	2564,86	−5101,16	−5095,84	−5095,46	−5102,09
7203.T	2744,62	2746,58	2748,77	2746,98	−5461,60	−5458,62	−5463,00	−5466,33
6758.T	2577,20	2576,27	2572,45	2574,81	−5126,78	−5118,00	−5110,36	−5121,99
9432.T	2622,33	2630,20	2614,81	2621,75	−5217,04	−5225,86	−5195,09	−5215,87

*Notes:*IGSV is the Inverse Gamma SV model, ASV is Asymetric SV model, GammaSV is the Gamma SV model, while the LNSV is the log normal SV model.

In terms of the log likelihood, the Inverse Gamma SV model is best model for the AAPL, AMZN and 6758.T. The asymetric SV model is the best for MSFT, 6861.T, and 9432.T. On the other hand the Log normal SV model is, better for NVDA and, 7203.T . Of these models the Inverse gamma SV model has the thicker tails. Taking into account the number of parameters using the BIC, the results are similar for the Inverse gamma SV model. However, the Log normal SV is the best 50% of the time, with the ASV model being substantially best for the 9432.T .

3.5 Conclusions

This paper obtained an analytic expression for the likelihood of an inverse gamma SV model. As a result, it is possible to obtain the Maximum Likelihood estimator. The exact value of the likelihood is also useful for Bayesian estimation and model comparison. Within the literature of nonlinear or non Gaussian state space models this novel approach is one of the very few methods that allow MLE because we are able to obtain the likelihood exactly. We provide the explicit formulas for this likelihood as well as the code to calculate it. Furthermore, we obtained the filtering and smoothing distributions for the inverse volatilities as mixture of gammas, allowing exact sampling from these distributions. Inverse gamma SV models can account for fat tails, which are observed in most

macroeconomic and financial data. The approach that we use to obtain the likelihood expression is a result of integrating out the volatilities in the model. This approach is computationally efficient, simple and accurate. The empirical fit of the inverse gamma SV model is better than other alternative models in the literature with inflation data for two countries and for 4 exchange rates series as shown in the empirical exercises. Lastly, using daily returns for 8 stocks on the Tokyo Stock Price Index (Topix) and the Standards and Poors 500 ($S\&P500$), the performance of the model is compared to the stochastic volatility model with leverage. The proposed model performs better 62.5% than the asymmetric SV model in terms of the Bayesian Information Criterion (BIC). The Inverse Gamma SV model can be extended to the multivariate case by using a factor model framework such as proposed in Kim et al. (1998), however, we leave the derivation of the exact likelihood for future research.

CHAPTER 4
COMMON INVERSE GAMMA STOCHASTIC VOLATILITY FACTOR IN VECTOR AUTOREGRESSIONS

4.1 Introduction

The literature on VARs with many outcome variables is receiving increased attention. Papers that have compared the structural analysis and forecasting performance of small VARs of less than 8 variables to that of large VARs with more than a dozen variables have often found these large VARs to outperform models with fewer outcome variables (e.g. Giannone et al. (2015), Carriero et al. (2016), Koop (2013) and Banbura et al. (n.d.)). The baseline models in the literature made the assumption that the variance is constant and the errors are gaussian. However, macroeconomic and financial data has been found to exhibit fat tails and time varying volatility. Thus, to introduce time variation in the volatility of the errors, macroeconomic and financial data literature has favoured the use of stochastic volatility models as they tend to provide better forecasts (e.g.Kim et al. (1998), C. Sims & Zha (2006) and Chan & Grant (2016a)). Empirical studies, such as Leon-Gonzalez (2018), Chiu et al. (2017) and Cross & Poon (2016) point out the improved forecasting performance of stochastic volatility models when fat tailed distributions are modelled in the error structure of macroeconomic variables.

Carriero et al. (2016) observed that the volatility estimate patterns across macroeconomic variables exhibit similarities across the variables in support of their common drifting volatility model. They employ a common factor model whose volatility is log normal such that the variance of the error structure is $var(e_t) = \sigma_t^2\Sigma$. A number of papers either apply or build on this specification of the variance to model this feature of macroeconomic variables (e.g Hartwig (2022), Mumtaz (2016), Poon (2018), Götz & Hauzenberger (2021) and Hou et al. (2023)) with the latter also including variations for

this Common Stochastic Volatility (CSV) model such as adding t innovations so that $var(e_t) = \lambda_t \sigma_t^2 \Sigma$, where λ_t is iid inverse gamma. Inclusion of this common volatility feature in the error structure consistently increased forecasting accuracy in the above literature.

Motivated by these findings, this paper proposes to improve the efficiency of CSV models by combining both properties of the heavy tailed distributions observed in most macroeconomic and financial applications, and the common factors. First, we obtain an analytical expression of the likelihood for the variance of the error structure such that $var(e_t) = \sigma_t^2 \Sigma$, where the time varying volatility in this case σ_t^2 is inverse gamma. This specification allows for heavy tails, which are an important feature of economic and financial data empirically. We show that by marginalising out the volatilities, the analytical expression of the likelihood that we obtain is simple to estimate and computationally efficient. Using this likelihood, we then obtain numerically, the marginal likelihood and the one step ahead out of sample forecasts that are the basis of comparison. We compare our method with related CSV models in the literature. We provide 2 applications for this comparison. The first application uses 20 macroeconomic variables each for Japan, Brazil, US, and the UK. Another application uses daily exchange rate returns for a small VAR of 4 currencies and a larger VAR of 8 currencies.

Section 4.2 reviews the literature. Section 4.3 describes the novel inverse gamma CSV model, derives the expression of the likelihood and provides some tests for a section of our algorithm. Section 4.4 details the models that will be used for comparison with our model and explains how we obtain the marginal likelihood and the average log predictive likelihood. In Section 4.5 we illustrate our novel algorithm to large data sets of quarterly macroeconomic data and financial data to compare the empirical performance of different models and finally section 4.6 concludes.

4.2 Literature Review

VARs are widely used in the structural analysis and forecasting literature ever since they were proposed by C. A. Sims (1980) because of their forecasting ability. Empirical analysis tends to use these VARs for model comparison as a benchmark for new approaches

because of their general simplicity. However, VARs tend to require the estimation of many parameters, thus, earlier analysis on Bayesian VAR (BVAR) methods such as Doan et al. (1984) typically analysed a small number of dependent variables. Bayesian methods are incorporated into VARs to allow for shrinkage by introducing informative priors. The forecasting performance for VARs of these small models of less than 10 dependent variables as a result was greatly improved (eg in Litterman (1986)).

Empirical macroeconomic applications however, tend to require an analysis of a very large number of dependent variables to reduce the omitted variable bias. Traditional approaches to estimating these large VARs, include panel VARs (e.g Poon (2018) and Canova et al. (2007)) or the factor models of (e.g Stock & Watson (2002) and Forni et al. (2000). The introduction of large BVARs that are unrestricted by Bańbura et al. (2010) provides an alternative approach to estimating a large number of variables. These large BVARs have been found to have better forecasting edge compared to smaller VARs (e.g Bańbura et al. (2010) Giannone et al. (2015) and Koop (2013)).

4.2.1 Minnesota Prior

Minnesota priors were first introduced by Doan et al. (1984) and further developed by Litterman (1986) as shrinkage priors for small VARs for auto regressive parameters. The basic principle behind the Minnesota prior comes from the belief that for a variable $y_{i,t}$, the coefficients of the lags closest to that variable e.g $y_{i,t-1}$, provide more accurate information compared to later lags $y_{i,t-2}, \ldots, y_{i,t-p}$. In addition, a variable's own lags have more information that can explain its own variable compared to other explanatory variable's lags. This prior belief has the effect of reducing the dimensionality problem arising from too many parameters associated with VARs significantly. The belief is expressed, by shrinking the values of the coefficients of later lags of own variables and all lags of other explanatory variables towards 0. Moreover, the diagonal elements of the coefficient of the earliest own lags can be shrunk towards 1. Shrinking the coefficients this way underpins the random walk preference for the Minnesota prior (Karlsson, 2013). Variants of this prior with specifications for stationary and persistent processes are also available.

To illustrate the Minnesota prior, we follow the specification in Bańbura et al. (2010):

$$Y_t = \mu + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + e_t$$

where e_t is normal distributed, $E(e_t e_t') = \Sigma$ and $\theta = (\theta_1', \ldots, \theta_p')'$ is a $n \times n$ matrix of coefficients of the lags of Y_t. This model has many parameters for the variance covariance matrix Σ and the matrix of coefficients, which can cause many problems for bigger dimensions of n. More specifically, Bayesian analysis of the Minnesota prior corresponds to imposing a prior on the parameters (μ, θ, Σ) as follows for the prior mean of the coefficients θ:

$$E[(\theta_p)_{ij}] = \begin{cases} 1, & \text{for } i = j, \text{ and p=1} \\ 0, & \text{otherwise} \end{cases}$$

Litterman applies a tighter shrinkage as p gets longer in line with prior beliefs expressed above for later lags of own variables. To apply a tighter shrinkage to lags of other explanatory variables, Litterman suggests setting the prior for the variance of a VAR with m variables as:

$$Var[(\theta_p)_{ij}] = \begin{cases} \frac{\pi_1^2}{(p^{\pi_3})^2}, & \text{for the coefficient of lag p of } y_i \text{ , where } i = (p-1)m + j \\ \frac{(\pi_1 \pi_2 s_j)^2}{(p^{\pi_3} s_r)^2}, & \text{for the coefficient of lag p of } y_{r \neq j}, \text{ where } i = (p-1)m + r \\ \infty, & \text{deterministic variables, } i = mp + 1, \ldots, k \end{cases}$$

where π_1^2 is the overall variance, π_2^2 is the variance for constant and deterministic components or relative tightness of other variables, π_3^2 is the rate of lag decay and $\frac{s_j^2}{s_r^2}$ accounts for the different variances, that is, s_j^2 is the OLS residual variance for equation j and s_r^2 is the equivalent for equation r. The prior mean for the deterministic components is usually set to 0. Alternatively, often times the very first lag is identified as an identity matrix.

The Minnesota prior is completed by imposing normal priors for the coefficients such

as:

$$vec(\theta) \backsim N(\bar{\theta}, \bar{\Omega} \otimes \Sigma)$$

where $\bar{\theta}$ includes intercepts, coefficients of a variables own lags and of lags of other explanatory variables, $\otimes$ is the Kronecker product and $\Sigma \backsim$ Inverse Wishart$(\bar{s}, \bar{v})$. For data in growth rates $\bar{\theta}$ is typically shrunk to 0 consistent with the prior belief that data in growth rates is not persistent, whereas, for data in levels $\bar{\theta}$ is shrunk to zero for all coefficients excluding the one for a variable's own lag that is shrunk to one also consistent with prior beliefs on the high persistence associated with data in levels.

The Minnesota prior sets $\bar{\Omega}$ as a diagonal matrix with diagonal elements depending on $\pi_1^2, \pi_2^2, \pi_3^2$ which are the key hyperparameters, thus reducing the prior elicitation process considerably. Litterman reviewed this prior by estimating the VAR equation by equation. Consequently, a number of variations to this prior can be found in the literature. For instance, Bańbura et al. (2010), Kadiyala & Karlsson (1997) and C. Sims & Zha (1998) modify the Minnesota prior so that the beliefs on the coefficients are symmetrical across all equations and thus make the posterior probability density function tractable with some restrictions.

4.2.2 Stochastic Volatility Models

To extend the literature on large BVARs, recent analysis introduce non linear time variation in the errors, which is a feature of macroeconomic and financial data literature. One way to do this is by using stochastic volatility models (e.g Kim et al. (1998), C. Sims & Zha (2006)). Early ideas of stochastic volatility were first proposed by Black (1976) for financial econometrics, and the first model for macroeconomics, using Bayesian methods was proposed by Jacquier et al. (1994). The model proposed by Jacquier et al. (1994) implied that the $var(e_t) = \sigma_t^2$ and estimation of the algorithm was achieved by sampling the volatilities individually. A drawback of this approach was the tendency to converge slowly as a result of the 'single-move' estimation of the volatilities.

Uhlig (1997), Shephard & Pitt (1997) and Cogley & Sargent (2001) among others, proposed to improve the convergence of these stochastic volatility models by sampling

the volatilities in blocks at a time. In an influential paper, Kim et al. (1998), show that the log volatilities of SV models can be marginalised out to estimate the unknown parameters. Both these approaches follow a log normal specification of the volatilities which implies that the distributions have thin tails (Madan & Seneta, 1990).

A key property of macroeconomic data is the heavy tails observed in their distributions. Empirical works, such as, Leon-Gonzalez (2018), Chiu et al. (2017) and Cross & Poon (2016) show that when fat tailed distributions are modelled in the error structure of macroeconomic variables these stochastic volatility models often provide better forecasts. Inverse gamma SV models imply student-t marginals, thus, they can account for fat tails that are observed in most macroeconomic and financial data (Leon-Gonzalez, 2018). Recent literature extends the stochastic volatility models by imposing varying assumptions on the error structure to improve evaluation and forecasting performance. We take a look at some of the specifications that have been proposed to model multivariate stochastic volatility models.

4.2.3 Common Stochastic Volatility Models

Carriero et al. (2016) observed that the volatility estimate patterns across macroeconomic variables, exhibit similarities across the variables. They proposed a common stochastic volatility approach to estimate stochastic volatility models. Including this common feature of macroeconomic data in the error structure, resulted in an increase in forecasting accuracy e.g Carriero et al. (2016). The results motivated the recent increase in specifying common stochastic volatility models whose specifications we now discuss in what follows.

The CSV literature capture the common factors using a log-normal process such that $\sigma_t^2 = \mathrm{e}^{h_t}$, where the law of motion of the log volatilities h_t, for $t = 2, \ldots, T$ is specified as a stationary $AR(1)$ process as follows:

$$h_t = \varphi h_{t-1} + \varepsilon_t^h, \qquad \varepsilon_t^h \sim N(0, \alpha^2)$$

and where $|\varphi| < 1$ and h_1 is normally distributed as $h_1 \sim N(0, \alpha^2/(1-\varphi^2)$.

Models such as in (Carriero et al. (2016), Hartwig (2022), Poon (2018) and Götz &

Hauzenberger (2021)) employ this specification of common factor model whose volatility is log normal such that the variance of the error structure is $var(e_t) = \sigma_t^2 \Sigma$. Specified in this way, the assumption is that the volatilities are proportional to each other given that they are all scaled by the same time varying parameter. This assumption though seemingly too restrictive, is supported nonetheless by empirical evidence on the co-movements observed in the volatility of macroeconomic variables e.g Carriero et al. (2018).

Adding t innovations to this CSV framework to account for heavy tailed distributions entails adding an additional hyper parameter for student t innovations so that we have $var(e_t) = \lambda_t e^{h_t} \Sigma$, where λ_t has an inverse gamma distribution with n degrees of freedom. $\lambda_1, \ldots, \lambda_t$ are assumed to be uncorrelated.

4.2.4 Moving Average Common Stochastic Volatility Models

The literature above assumes, mostly for computational convenience that the error terms are independent from each other. However, the error structure of time series variables are known to be serially correlated as illustrated in the seminal paper by Slutzky (1937). Early applications of the dependency in the error structure referred to this as the Moving Average structure. Chan (2013) models this property in a univariate moving average stochastic volatility model by assuming that the errors in the measurement equation are serially dependent. The framework for this approach to stochastic volatility models is as follows:

$$\begin{aligned} y_t &= \mu_t + e_t \\ e_t &= u_t + \psi_1 u_{t-1} + \ldots \psi_q u_{t-q} \qquad & u_t &\sim N\big(0, e^{h_t}\big) \\ h_t &= \mu_h + \phi_h(h_{t-1} - \mu_h) + \eta_t \qquad & \eta_t &\sim N\big(0, \sigma_h^2\big) \end{aligned} \tag{4.2.1}$$

where q is the lag order of the moving average process and u_t and η_t are assumed to be uncorrelated. Both ψ_1 and ϕ_h are less than 1 to satisfy invertibility conditions. Chan (2013) found in their application to inflation that their model that specified the moving average structure in stochastic volatility models resulted in a better fit compared to the models that assume serial independence.

Serially dependent errors can be added to the general framework of the common

stochastic volatility models to provide richer dynamics of the error structure. Chan (2020) propose a framework that combines the MA structure in (4.2.1) and falls within the CSV framework proposed by Carriero et al. (2016) such that the variance of the error structure is log normal as before, that is $var(e_t) = \sigma_t^2 \Sigma$. Assuming that e_t is described by a moving average process with 1 lag MA(1), that is

$$e_t = u_t + \psi_1 u_{t-1}$$

Then the covariance matrix Ω structure would be as in Chan (2020) as follows:

$$\Omega = \begin{pmatrix} (1+\psi_1^2)\mathrm{e}^{h_1} & \psi_1 \mathrm{e}^{h_1} & 0 & \cdots & 0 \\ \psi_1 \mathrm{e}^{h_1} & \psi_1^2 \mathrm{e}^{h_1} + \mathrm{e}^{h_2} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \psi_1^2 \mathrm{e}^{h_{T-2}} + \mathrm{e}^{h_{T-1}} & \psi_1 \mathrm{e}^{h_{T-1}} \\ 0 & \cdots & 0 & \psi_1 \mathrm{e}^{h_{T-1}} & (1+\psi_1^2)\mathrm{e}^{h_T} \end{pmatrix} \tag{4.2.2}$$

Adding student-t innovations would result in the model having an additional hyper parameter as before and thus the diagonal of the covariance matrix would be multiplied by the hyper parameter $\lambda_{1:t}$.

4.2.5 Testing MCMC Samplers

Occasionally when coding MCMC samplers, there are bound to be errors. The errors could be in the algorithm itself in which case it may be incorrect mathematically or in other instances the algorithm may fail to converge to the correct solution. To check errors in implementing the code for these posterior simulators, that is, the mathematical correctness of the MCMC algorithms Geweke (2004) provides a test that is often referred to as the Geweke test. This test has proved to be one of the preferred standards for checking and testing errors in MCMC samplers.

The test exploits a very simple idea that supposing one wishes to test the MCMC posterior simulator for a generative model with data denoted by Y and parameters denoted

as θ for the posterior $\pi(\theta|Y)$. There are two ways that can be used to draw from the joint distribution $\pi(\theta, Y)$. One way is to draw M values of θ from the prior, $\theta_1, \ldots, \theta_M$ and calculate some statistic for the moments $X = f(\theta_1, \ldots, \theta_M)$. An alternative way is to sequentially, for $i = 1, \ldots, M$ do as follows:

Step 1. Draw θ_i from its prior distribution $\pi(\Theta)$
Step 2. Draw Y_i from the conditional distribution $\pi(Y|\theta_i)$
Step 3. Use MCMC sampler to obtain $\tilde{\theta}_i$ from the posterior $\pi(\theta|Y = Y_i)$

Then compute the moments $Y = f(\tilde{\theta}_1, \ldots, \tilde{\theta}_M)$. Either draws will preserve the samples as draws from the stationary distribution such that its like sampling from the joint distribution $\pi(\theta, Y)$. If there is no error in the MCMC Sampler, then if one compares the moments X and Y the difference will not be statistically significant. Geweke (2004) proposes formal frequentist hypothesis tests to check whether the joint distributions are significantly different. Geweke (2004) does so by comparing, moments of the distributions such as the mean and the standard deviation. One way to compare the moments of the two distributions can be implemented using Z-tests under the null hypothesis that the MCMC sampler is correct as follows:

$$Z = \frac{X - Y}{\sqrt{Variance\ of\ X + Variance\ of\ Y} * \sqrt{1/n}}$$

We will use this test to check our implementation of the code that we use to draw the unknown time varying volatilities.

4.3 Inverse Gamma CSV Model

The model can be described as follows:

$$Y_t = x_t\beta + e_t, \qquad \text{with} e_t|x_t \sim N\left(0, \sigma_t^2\Sigma\right), \qquad t = 1, \ldots, T, \qquad (4.3.1)$$

where Y_t is a $r \times 1$ vector, β is a matrix of unknown parameters, x_t is a vector of predetermined variables, and e_t is a $r \times 1$ vector of unobserved errors .

Define the time varying stochastic process as $k_t = (\sigma_t^2)^{-1}$, and assume that $k_t = z_t' z_t$, where z_t is an $n \times 1$ vector. The vector z_t has the following Gaussian AR(1) representation:

$$z_t = z_{t-1}\rho + \epsilon_t \qquad\qquad vec(\epsilon_t) \sim N(0, \theta^2 I_n)$$

The parameter ρ controls the persistence of the volatility and n represents the degrees of freedom of the non central chi squared distribution and this parameter will be estimated. As a normalization we assume $\theta^2 = 1$ so that we have two volatility parameters to estimate. This representation of z_t implies that the conditional distribution of $k_t|k_{t-1}$ is a non central chi squared. The non central chi-squared distribution is well defined for non integer values of n, therefore we will treat the unknown parameter n as continuous. Given the properties of a gamma, the conditional mean of the inverse time varying volatility k_t, given its previous history is a weighted average of the unconditional mean of k_t and its previous value k_{t-1} as follows:

$$E(k_t|k_{t-1}) = \rho^2 k_{t-1} + (1 - \rho^2)E(k_t)$$

where $\rho^2 k_{t-1}$ represents the non centrality parameter. As is characteristic of time varying variance models, k_t is correlated with its previous value. Assuming that k_t is drawn from a stationary distribution with n degrees of freedom and using the properties of a gamma distribution, we have that the stationary and initial distribution of k_1 is:

$$k_1 \backsim Gamma\left(\frac{n}{2}, \frac{2}{1-\rho^2}\right)$$

To ensure stationarity, a necessary condition is $|\rho| < 1$. Thus, this model falls within the same framework as in the CSV literature that follows the seminal paper of Carriero et al. (2016) in that only σ_t^2 varies with time. In this model however, σ_t^2 is inverse gamma whereas the CSV literature has σ_t^2 following a log normal distribution. The inverse gamma specification implies a student-t distribution for y_t thus enabling us to model heavy tailed distributions. By integrating out the volatilities analytically, we obtain analytical expressions of the likelihood that are simple as follows:

4.3.1 Likelihood

Define $\varepsilon_t^2 = e_t' \Sigma^{-1} e_t$, then, the conditional likelihood function given the gaussian assumption on e_t, is as follows:

$$L(Y_1, \ldots, Y_T \mid \Sigma, k_{1:T}) = \left(\frac{1}{(2\pi)^{\frac{r}{2}}}\right)^T \left(\prod_{t=1}^{T} |\Sigma|^{-\frac{1}{2}}\right) \left(\prod_{t=1}^{T} |k_t^{-1}|^{-\frac{r}{2}}\right) \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \varepsilon_t^2 k_t\right) \tag{4.3.2}$$

$k_{1:T}$ however, is not known, so we marginalise it out to obtain the likelihood conditional on Σ , that is, $L(Y_1, \ldots, Y_T|\Sigma)$. Thus, marginal on the unknown time varying volatility k_t, the expressions of the likelihood are as in the proposition below, the proof of which is in the Appendix:

Proposition 4.3.1.

$$L(Y_1) = (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} 2^{\frac{r}{2}} \frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left|\varepsilon_1^2 + V_1^{-1}\right|^{-\frac{n+r}{2}} V_1^{-\frac{n}{2}}$$

for the second is:

$$L(Y_2|Y_1) = (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} \frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}} \frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{(\varepsilon_2^2 + 1)^{-\frac{n+r}{2}}}{(1-\delta_2)^{-\frac{n+r}{2}}} \hat{C}_2$$

for the third is:

$$L(Y_3|Y_2, Y_1) = (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} \frac{1}{c_3} \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \frac{\Gamma\left(\frac{n+r+2h_2}{2}\right)}{(\varepsilon_3^2 + 1)^{\frac{n+r}{2}}} (2S_3)^{\frac{n+r+2h_2}{2}} \frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}} \frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \hat{C}_3$$

for the fourth is:

$$L(Y_4|Y_3, Y_2, Y_1) = (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} \frac{1}{c_4} \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} \frac{\Gamma\left(\frac{n+r+2h_3}{2}\right)}{(\varepsilon_4^2 + 1)^{\frac{n+r}{2}}} (2S_4)^{\frac{n+r+2h_3}{2}} \frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}} \frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \hat{C}_4$$

and for any $t \geq 3$ is

$$L(Y_t|Y_{1:t-1}) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_t}\sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}\frac{\Gamma\left(\frac{n+r+2h_{t-1}}{2}\right)}{(\varepsilon_t^2+1)^{\frac{n+r}{2}}}(2S_t)^{\frac{n+r+2h_{t-1}}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_t$$

where:

$$V_1 = (1-\rho^2)^{-1}$$
$$\widetilde{V}_2^{-1} = V_1^{-1} + \varepsilon_1^2$$
$$\delta_2 = \rho^2(\widetilde{V}_2^{-1} + \rho^2)^{-1}$$
$$Z_2 = (\varepsilon_t^2+1)^{-1}\delta_2$$
$$\widetilde{C}_{2,h_2} = \frac{[(n+r)/2]_{h_2}}{[n/2]_{h_2}}\left(\frac{1}{2}\rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}\right)^{h_2}\frac{1}{h_2!}$$
$$\widetilde{C}_{3,h_3} = \sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+r+2h_2}{2}\right)\frac{[(n+r)/2+h_2]_{h_3}}{[n/2]_{h_3}}\left(\frac{1}{2}\rho^2 S_3\right)^{h_3}\frac{1}{h_3!}(2S_3)^{\frac{n+r+2h_2}{2}}$$
$$c_3 = {}_2F_1\left(\frac{n+r}{2},\frac{n+r}{2};\frac{n}{2};\delta_3\right)\Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$
$$\hat{C}_t = {}_2F_1\left(\frac{n+r+2h_{t-1}}{2},\frac{n+r}{2};\frac{n}{2};Z_t\right) \text{ for } t \geq 2 \text{ and where } h_1 = 0$$

for $T \geq t \geq 3$

$$S_t = (\varepsilon_{t-1}^2 + 1 + \rho^2)^{-1}$$
$$\widetilde{V}_t^{-1} = \varepsilon_{t-1}^2 + 1$$
$$Z_t = (\varepsilon_t^2+1)^{-1}S_t\rho^2$$
$$\delta_t = \left((1-\rho^2 S_t)^{-1}S_t\rho^2(\widetilde{V}_{t-1}^{-1}+\rho^2)^{-1}\right)$$

and for $T+1 \geq t \geq 4$

$$c_t = \sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}(1-\rho^2 S_t)^{-\frac{n+r+2h_{t-1}}{2}}\Gamma\left(\frac{n+r+2h_{t-1}}{2}\right)(2S_t)^{\frac{n+r+2h_{t-1}}{2}}$$

$$\widetilde{C}_{t-1,h_{t-1}} =$$
$$\sum_{h_{t-2}=0}^{\infty}\widetilde{C}_{t-2,h_{t-2}}\Gamma\left(\frac{n+r+2h_{t-2}}{2}\right)\frac{[(n+r)/2+h_{t-2}]_{h_{t-1}}}{[n/2]_{h_{t-1}}}\left(\frac{1}{2}\rho^2 S_{t-1}\right)^{h_{t-1}}\frac{(2S_{t-1})^{\frac{n+r+2h_{t-2}}{2}}}{h_{t-1}!}$$

and $S_{T+1} = (1 + \varepsilon_T)^{-1}$

$[x]_h$ denotes the rising factorial and ${}_2F_1$ a gauss hypergeometric function (e.g. Muirhead (2005, p. 20)). These hypergeometric functions can be transformed to accelerate their convergence in a number of ways. Abramowitz et al. (1988, p.559) defines several transformations such as the Euler transformation where:

$$ {}_2F_1(a, b; c; z) = (1 - z)^{c-a-b} {}_2F_1(c - a, c - b; c; z) $$

or a linear combination approach:

$$ \begin{aligned} {}_2F_1(a, b; c; z) = & \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)} {}_2F_1(a, b; a + b - c + 1; 1 - z) \\ & + (1 - z)^{c-a-b} \frac{\Gamma(c)\Gamma(a + b - c)}{\Gamma(a)\Gamma(b)} {}_2F_1(c - a, c - b; c - a - b + 1; 1 - z) \\ & \text{for } (|arg(1 - z)| < \pi) \end{aligned} $$

$\hat{C}_t$ above transformed using the Euler transformation is thus:

$$ \hat{C}_t = (1 - Z_t)^{-\frac{n+2r+2h_{t-1}}{2}} {}_2F_1\left(-\frac{r + 2h_{t-1}}{2}, -\frac{r}{2}; \frac{n}{2}; Z_t \right) \text{ for t} \geq 2 \text{ and where } h_1 = 0 $$

However, in our coding we used the Euler acceleration only for $\hat{C}_2$ and c_3. Instead, we accelerated the calculations by implementing parallel computing in the code. This is possible because many of the coefficients in the series are the same for every t, therefore they only need to be computed once, which can be done in parallel. We also calculate all the $\hat{C}_t$ in parallel.

4.3.2 Geweke's test

To test the implementation of the algorithm that we use to draw the unknown time varying volatilities $k_{1:t}$ for this approach, we use the Geweke test described above as follows:

Step 1. Draw $k_{1:t}$ *from a gamma distribution which is our prior for* $\pi(k_{1:t})$

Step 2. Draw the data from the normal distribution $\pi(Y_{1:t}|k_{1:t})$

Step 3. Draw $k_{1:t}|Y_{1:t}$ *using the posterior simulator*

This is done repeatedly. If our implementation of this algorithm is correct, we expect by checking the simulation that the distribution of $k_{1:t}$ should be the same in step 1 and in step 3. We use the Z-test to test whether the means for the time varying volatility $k_{1:t}$ drawn from the prior in step 1 (labelled as X in 4.1) are significantly different from the ones drawn using the posterior simulator in step 3 $k_{1:t}$ (labelled as Y). We obtain the following Z statistics, which under the null hypothesis that the algorithm is correct follow $Z \backsim N(0,1)$:

Table 4.1: Z test statistics for the mean

	X	**Y**
Standard Deviation	19.56542	19.91627
Mean	36.79686	36.65944
Mean Difference = 0.17169		
Critical value (two tails) at 0.05 significance level =1.96		
Z= 0.95268		

The Z score is less than the critical value corresponding to the 0.05 significance level of 1.96, thus we fail to reject the null hypothesis that the algorithm implementation is correct.

4.4 Models for Comparison

We compare our proposed model to 5 model specifications to demonstrate its merits. First we compare with a plain homoscedastic BVAR. We then compare with four other CSV models, that is, Carriero et al. (2016)'s specification of heteroscedastic innovations. As Hou et al. (2023) and Chan (2020), have done in these applications we label the model a BVAR-CSV and its variance structure is such that $var(e_t) = \sigma_t^2\Sigma$. The third model adds t innovations to this specification such that $var(e_t) = \lambda_t\sigma_t^2\Sigma$, where λ_t is iid inverse

gamma. This is the BVAR-CSV-t model. The fourth model is the BVAR-CSV-MA model that adds moving average (MA)innovations to the error such that $e_t = \epsilon_t + \psi\epsilon_{t-1}$. We add t innovations to this MA specification to get the BVAR-CSV-MA-t model. All these CSV models have the same description as in (4.3.1) with the exception of the variance of the error term. These models have the advantage of having a kronecker product in the structure of their covariance which speeds up computations significantly in large models. In comparison to our proposed model, the BVAR-CSV model has the same number of parameters. The only difference is the change of distribution. Thus, the comparison of the two models adds to the literature on the merits or lack thereof, of including the property of heavier tailed distributions observed in macroeconomic and financial data. Table 4.2 lists the comparison models.

Table 4.2: Models for Comparison

Model	**Code**	**Description**
BVAR	M_1	Standard BVAR with homoscedastic errors
BVAR-CSV	M_2	BVAR with a CSV
BVAR-CSV-t	M_3	BVAR with both CSV and t innovations
BVAR-CSV-MA	M_4	BVAR with CSV and MA(1) innovations
BVAR-CSV-MA-t	M_5	BVAR with CSV, MA(1) and t innovations
BVAR-CSV-IG	M_6	BVAR with CSV and inverse gamma innovations

The priors for those parameters that are common to all models are similar for comparison purposes.

4.4.1 Comparison Methods

We use the marginal likelihood and predictive likelihood to compare the models. The closed form expression of the marginal likelihood for the homoscedastic BVAR with a natural conjugate prior is as proposed by Karlsson (2013). The expressions of the marginal likelihood for the other models under comparison can be obtained numerically by following the method proposed by Chib (1995), which is known as the Chib's method. This is

the method that we use for all the models, except for the inverse gamma CSV, for which we use importance sampling. Chib's method uses the Bayes Theorem as follows:

$$\pi(\theta|Y) = \frac{\pi(\theta)\pi(Y|\theta)}{\pi(Y)} \tag{4.4.1}$$

where $\pi(\theta|Y)$ is the posterior of the vector of parameters θ given Y, the numerator is the prior for the parameters multiplied by the likelihood and the denominator is the marginal likelihood. For a large class of state space models $\pi(Y|\theta)$ is not known, thus numerical methods such as particle filters or importance sampling are used to approximate it. For the inverse gamma with $\Sigma_t = \Sigma$, we have the analytical expressions of the likelihood, thus we do not need numerical methods. Then, given the likelihood, (4.4.1) can be rearranged to obtain:

$$\pi(Y) = \frac{\pi(\theta)\pi(Y|\theta)}{\pi(\theta|Y)}$$

The log marginal likelihood can thus be obtained as:

$$log(\pi(Y)) = log(\pi(\hat{\theta})) + log(\pi(Y|\hat{\theta})) - log(\pi(\hat{\theta}|Y))$$

where $log(\pi(Y|\hat{\theta}))$ is the log likelihood obtained above for some value of θ, e.g the posterior mean $\hat{\theta} = E(\theta|Y)$.

To implement Chib's method, Chan (2020) does so by evaluating the integrated likelihood marginal on the state variables. To evaluate the marginal likelihood in this way is itself computationally challenging, more so for more complex models as it involves high dimensional integration. Let z represent the state variables, then the integrated likelihood proposed in Chan (2020) can be obtained as follows:

$$\pi(Y|\theta) = \int \pi(Y|\theta, z)\pi(z|\theta)dz \tag{4.4.2}$$

where $\pi(z|\theta)$ is the prior density of the latent variables conditional on θ. For the BVAR models with a common stochastic volatility specification and t innovations in the comparison models above, this integrated likelihood can be evaluated numerically by marginalising out all the state variables in z. Chan (2020) uses an importance sampling

approach proposed in Chan & Grant (2016b). Details of the approach proposed in this paper to evaluate the integrated likelihood for each of the models above can be obtained from the online appendix of their paper.

Importance sampling in general works well provided that a "good" approximation to the posterior density $\hat{\pi}(z|Y,\theta)$ is known. This importance sampling density should be such that $\pi(Y|\theta,z)\pi(z|\theta)$ is dominated by it. Then, (4.4.2) can be written as follows:

$$\pi(Y|\theta) = \int \pi(Y|\theta,z)\pi(z|\theta)dz = \int \frac{\pi(Y|\theta,z)\pi(z|\theta)\hat{\pi}(z|Y,\theta)}{\hat{\pi}(z|Y,\theta)}dz = \int W(z(\hat{\pi}(z|Y,\theta)dz$$

where $\hat{\pi}(z|Y,\theta)$ is the importance density, and $W(z) = \frac{\pi(Y|\theta,z)\pi(z|\theta)}{\hat{\pi}(z|Y,\theta)}$ is the importance weight which ideally should have a small variance.

Consequently, if $z_1,\ldots,z_N \backsim_{iid} \hat{\pi}(z|Y,\theta)$, then this integral is approximately equal to averaging the ratio $\sum_{i=1}^{M} W_i(z_i)$ where W_i are the weights representing the ratio of densities (Kroese et al., 2013). Therefore, for this approach to work, the weight has to have a small variance, meaning that the importance sampling density $\hat{\pi}(z|Y,\theta)$ has to be similar to the conditional density of the latent variables $\pi(z|Y,\theta)$, which may be difficult in big dimensions.

Obtaining the marginal likelihood for our model is simpler, because we have an exact analytical expression for the integrated likelihood. We use an importance sampling approach. We assume as follows, given our parameters $\tilde{\theta} = (\theta, n, \rho)$, where $\theta = \beta, \Sigma$. Let M_1 be a VAR model with a time varying variance such that var(M_1)=$\Sigma\hat{\sigma}_t$. We further assume that $\hat{\sigma}_t$ is known, thus, for this model the marginal likelihood is known. Let M_2 be our CSV-IG model, in which $\hat{f}(n,\rho)$ is an approximation of the conditional posterior $\pi(n,\rho|Y,M_2)$. Because the prior $\pi(n|M_2) \sim$ is a log normal, we choose $\hat{f}(n)$ to be a log normal, with mean and variance from the posterior. Similarly, because the prior $\pi(\rho|M_2)$ is a beta distribution, we specify $\Rightarrow \hat{f}(\rho)$ as a beta distribution with parameters chosen to match the posterior mean and variance of ρ.

Thus, we can approximate the marginal likelihood by first approximating the following Bayes factor:

$$\frac{\pi(Y|M_1)}{\pi(Y|M_2)} = \int \frac{\pi(\theta|M_1)\pi(Y|\theta, M_1)\hat{f}(n,\rho)}{\pi(\theta|M_2)\pi(Y|\tilde{\theta}, M_2)\pi(n,\rho|M_2)}\pi(\tilde{\theta}|Y, M_2)d\tilde{\theta}$$

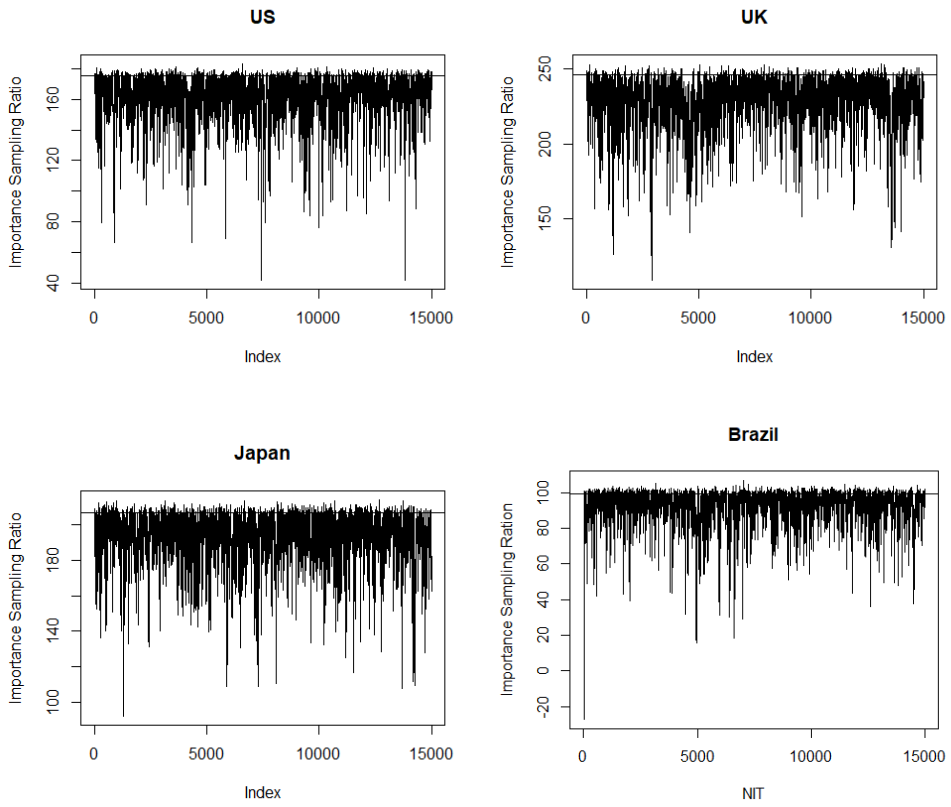
This integral can be calculated by importance sampling, where the weight for each $\tilde{\theta}$ is thus defined as

$$W(\tilde{\theta}_i) = \frac{\pi(\theta_i|M_1)\pi(Y|\theta_i, M_1)\hat{f}(n_i,\rho_i)}{\pi(\theta_i|M_2)\pi(Y|\tilde{\theta}_i, M_2)\pi(n_i,\rho_i|M_2)}$$

where $\tilde{\theta}_i = (\theta_i, n_i, \rho_i)$ are obtained with the MCMC sampler for M_2. Thus, the Bayes Factor can be approximated with $\frac{1}{M}\sum_{i=1}^{M} W(\tilde{\theta}_i)$, where M is the number of random draws from the posterior.

Figure 4.1 shows the importance sampling ratios obtained from 15000 iterations with a burn in of 1000 of the sampler using our approach for the macroeconomic data. The horizontal line indicates the estimated value of the Bayes factor. Approximately 5% of the weights go beyond the horizontal line indicating good performance.

Figure 4.1: Importance Sampling Ratios



Importance Sampling Ratios obtained from 15000 iterations.

We also evaluate the performance of the models by using predictive likelihoods which measure the one step ahead out of sample forecasting accuracy. We provide one step ahead out of sample density forecasts computed from the predictive density. To evaluate the one step ahead density forecasts, a popular metric used is the predictive likelihood. The predictive likelihood, can be obtained from the marginal likelihood as follows:

$$Predictive\ Likelihood_{T_0+1:T} = \frac{Marginal\ Likelihood_{1:T}}{Marginal\ Likelihood_{1:T_0}}$$

Taking logs we can obtain the log predictive likelihood as:

$$log(\pi(Y_{(T_0+1):T})|Y_{1:T_0}) = log(\pi(Y_{1:T})) - log(\pi(Y_{1:T_0}))$$

One way to evaluate this log predictive likelihood is to obtain the log marginal likelihoods given data to the time periods above and obtain the difference. Another approach

is to estimate the model repeatedly, that is, estimate the model using data for $t = T_0$ and then forecast for $T_0 + 1$ and obtain $\pi(Y_{T_0+1}|Y_{1:T_0}) = \sum_{i=1}^{M} \frac{\pi(Y_{T_0+1}|\theta_i, Y_{1:T_0})}{M}$. Then move one step forward and repeat the process using data up to $T_0 + 1$ and forecast for $T_0 + 2$, and so forth. We use the former approach.

The Average Log Predictive Likelihood (ALPL), can be obtained by averaging over the number of periods, that is:

$$ALPL = \frac{log(\pi(Y_{(T_0+1):T})|Y_{1:T_0})}{T - T_0}$$

A larger ALPL implies better forecasting accuracy compared to the benchmark. Whenever the prior is based on the data, for example in the Minnesota prior, we use only data up to T_0 to train the prior in all cases.

4.5 Empirical Application

To illustrate the efficiency and usefulness of our proposed model addition to the CSV literature, we provide 2 applications. The first application uses 20 macroeconomic and financial variables each for Japan, Brazil, US, and the UK. Another application uses daily exchange rate returns for a small VAR of 4 currencies and a larger VAR of 8 currencies.

4.5.1 Macroeconomic Application

Vintage US macroeconomic data for the empirical application was obtained from the Federal Reserve Bank of Philadelphia while the financial variables were sourced from the Federal Reserve Bank of St Louis. For consistency with previous literature on CSV applications, these variables are the same as those used in e.g Chan (2020), Carriero et al. (2016) and Koop (2013) updated to 2022Q2. The variables include Real output, personal consumption expenditures, Investments, federal interest rates and the S&P500.

Variables for Japan were obtained from the Federal Reserve Bank of St Louis with the exception of three variables that were obtained from CEIC data. That is, the foreign effective exchange rate and the monetary base cited by CEIC as sourced from the Bank of Japan while the industrial production index was obtained from the International Mon-

etary Fund. All variables were chosen to closely match the 20 US variables, as such, the index of aggregate weekly hours for Japan represent hourly earnings for manufacturing whereas housing starts are obtained as data for work started on construction, dwellings or residential buildings as a total.

UK variables were obtained from the Federal Reserve Bank of St Louis. Long term government 10 year bond yields replace the 10 year treasury constant maturity rate. The Import price index for the UK is for all goods and services classified by origin. The variables for Brazil were also obtained from the Federal Reserve Bank of St Louis with the exception of 7 variables obtained from CEIC data sourced from various sources. The industrial production index, producer price index and the payroll index was cited as sourced from the Brazilian Institute of Geography and Statistics. The import price index was sourced from the Centre for Foreign Trade Studies Foundation. Government bond yields were sourced from the National Treasury Secretariat. The monetary base was sourced from the Central Bank of Brazil. Lastly, the Equity Market Index Sao Paulo Stock Exchange was calculated from the daily BOVESPA index.

Monthly data is converted to quarterly observations by obtaining their 3 monthly average values for the corresponding quarter. The comprehensive list with descriptions for the variables and the transformation employed is listed in Table 4.3.

Table 4.3: Variables Description

Variables Description	**Transformation**	**US**	**UK**	**JP**	**BR**
Real GNP/GDP	400 Δ log	○	○	○	○
Real Personal Consumption Expenditure	400 Δ log	○	○	○	○
Real Gross Private Domestic Investments:Nonresidential	400 Δ log	○			
Real Gross Private Domestic Investments:Residential	400 Δ log	○	○	○	
Real Net Exports of Goods and Services	None	○	○	○	○
Nominal Personal Income	400 Δ log	○			
Industrial Production Index	400 Δ log	○	○	○	○
Unemployment Rate	None	○	○	○	○
Nonfarm Payroll Employment	400 Δ log	○			○
Indexes of Aggregate Weekly Hours:Total	400 Δ log	○	○	○	
Housing Starts	400 Δ log	○	○	○	
Price Index for Personal Consumption Expenditures, Constructed	400 Δ log	○		○	○
Price Index for Imports of Goods and Services	400 Δ log	○	○		○
Effective Federal Funds Rate	None	○			○
1 Year Treasury Constant Maturity Rate	None	○	○		
10 Year Treasury Constant Maturity Rate	None	○	○		○
Moody's Seasoned Baa Corporate Bond Minus Federel Funds Rate	None	○			
ISM Manufacturing PMI Composite Index	None	○			
ISM Manufacturing New Orders Index	None	○			
S&P500	400 Δ log	○			
Producer Production Index	400 Δ log		○	○	○
Consumer Price Index	400 Δ log		○	○	○
Interest Rates ,Government Securities, Government Bonds	None		○	○	
Spot Exchange Rates	400 Δ log		○	○	○
M1	400 Δ log		○	○	○
M2	400 Δ log		○	○	○
Foreign Effective Exchange Rate	400 Δ log		○	○	○
Total Share Prices for All Shares	400 Δ log		○	○	○
Basic Discount Rate	None		○	○	○
Monetary Base	400 Δ log			○	○
Nikkei225	400 Δ log			○	
Equity Market Index Sao Paulo Stock Exchange	400 Δ log				○

4.5.2 Estimation Results

We use marginal likelihoods and the one step ahead, out of sample average log predictive likelihoods to compare the models using the data variables in Table 4.2 for the 4 countries. The ALPL are obtained as the difference in marginal likelihoods $log(\pi(Y_{1:T}) - log(\pi(Y_{1:T_0})$ divided by $T - T_0$ where T_0 represents observations for the 4th quarter of 2015. A larger value of the ALPL implies better forecasting performance. We use two types of priors for

Σ for all model estimations for easy comparison. The first prior is centered in the identity matrix where the prior mean $E(\Sigma) = I$ with $r + 3$ degrees of freedom. The second prior has the prior mean estimated as $E(\Sigma) = \hat{\Sigma}_{OLS}$, where $\hat{\Sigma}_{OLS}$ is a diagonal matrix estimated with OLS residuals. In both cases, the prior for the slope coefficients is of Minnesota type. The hyperparameters are set to be $\pi_1 = 0.04$ and $\pi_2 = 10^2$ following related literature. Table 4.4 shows the values of the Marginal likelihood for the BVAR-CSV-IG Model and the related standard errors.

Table 4.4: BVAR-CSV-IG Model Marginal Likelihoods and Standard Errors

Prior	**Evaluation Period**	Marginal Likelihoods(Standard Errors)			
		UK	**Japan**	**US**	**Brazil**
$E(\Sigma) = I$	24	-9325.0	-10693.5	-11826.8	-7333.9
		(0.1065)	(0.1319)	(0.1132)	(0.0446)
	50	-9334.2	-10668.3	-11786.4	-7335.6
		(0.0933)	(0.0933)	(0.1203)	(0.0540)
$E(\Sigma) = \hat{\Sigma}_{OLS}$	24	-8926.9	-10318.6	-11436.9	-6980.8
		(0.0932)	(0.1110)	(0.0638)	(0.1369)
	50	-8937.0	-10339.4	-11434.3	-6990.8
		(0.0942)	(0.1381)	(0.1147)	(0.1094)

The marginal likelihoods have a standard error which is smaller than 1 in all cases. Thus, when you divide by the number of observations in the evaluation period, then the standard errors become very small.

The marginal likelihood estimates in Table 4.5 for model 2 to 5 are obtained from 150000 simulations with 20000 burns. A larger value of this comparison metric implies that the observed data is more likely under that model. Following related literature the lag length is set to 4. The model with the better marginal likelihood and ALPL is indicated in double asterisks while the second best model has one asterisk.

Table 4.5: Marginal Likelihood and ALPL for $T - T_0 = 24$, $E(\Sigma) = I$

	Marginal Likelihood				ALPL			
Model	**US**	**UK**	**Japan**	**Brazil**	**US**	**UK**	**Japan**	**Brazil**
M_1	−9916,9	−11 351,1	−12 229,1	−8062,9	−57,92	−59,67	−44,53	−71,76
M_2	−9367,7	−10 702,4	−11 836,1	−7330	−42,92	−47,29	−42,88*	−71,34
M_3	−9351,17	−10 679,2**	−11 830,2*	−7323,8*	−42,55*	−47,62	−43,02	−71,17*
M_4	−9297,5*	−10 707,1	−11 840,8	−7328,9	−43,08	−47,28*	−42,9	−71,32
M_5	−9279,3**	−10 681,8*	−11 834,6	−7328,2*	−42,63	−47,55	−43,21	−71,18
M_6	−9325,0	−10 693,5	−11 826,8**	−7333,9	−41,47**	−47,15**	−42,83**	−71,08**

Notes: The best model is marked by ** and the second best by *.

Comparing the marginal likelihoods, heavy tailed distribution models with t innovations, that is, the BVAR-CSV-MA-t, the BVAR-CSV-t and BVAR-CSV-IG consistently outperform their complementary models with gaussian errors, thus reinforcing the evidence for error structures with t innovations. This is consistent with e.g Chiu et al. (2017) and Chan (2020) who also find that allowing for heavy tails improves model fit. A known drawback of the marginal likelihood is its sensitivity to the prior distribution, more so when t is small ((Gelman et al., 2013) and (Chan & Grant, 2016b)). The ALPL on the other hand is less sensitive to the prior distribution. The table shows that the BVAR-CSV-IG model has a better fit among all the CSV models under comparison thus, this added feature of Inverse Gamma innovations also improves the density forecasts.

Table 4.6 shows the marginal likelihood and ALPL comparisons for $T - T_0 = 50$ to allow for a longer evaluation period.

Table 4.6: Marginal Likelihood and ALPL for $T - T_0 = 50$, $E(\Sigma) = I$

	Marginal Likelihood				ALPL			
Model	**US**	**UK**	**Japan**	**Brazil**	**US**	**UK**	**Japan**	**Brazil**
M_1	−9923,17	−11 360,91	−12 232,17	−8076,98	−45,03	−50,85	−42,51	−56,70
M_2	−9378,34	−10 716,44	−11 844,52	−7331,87	−37,67	−43,54*	−41,41**	−67,89
M_3	−9362,01	−10 689,85*	−11 837,01	−7326,00**	−37,47	−43,63	−41,50*	−67,83*
M_4	−9305,53*	−10 721,28	−11 847,73	−7330,7	−37,54	−43,57	−41,58	−67,92
M_5	−9288,22**	−10 694,24	−11 835,46*	−7330,39*	−37,32*	−43,96	−41,60	−67,86
M_6	−9334,2	−10 668,3**	−11 786,4**	−7335,6	−36,96**	−43,44**	−42,19	−67,78**

Notes: The best model is marked by ** and the second best by *.

When the cut off point is increased to 50 periods, the BVAR-CSV-IG model has the

best fit for US, UK and Brazil data while the BVAR-CSV has the best fit for Japan data in terms of the average log predictive likelihood.

We compare the models using an alternative prior for the var-cov matrix that is centered on previous OLS estimates of the variance in the Minnesota prior style. The prior mean for Σ is thus a diagonal matrix, where the diagonal is estimated using OLS residuals with observations up to T_0, that is, $E(\Sigma) = \hat{\Sigma}_{OLS}$ where $\Sigma \backsim \text{IW}(S_0, df)$. Such that the expected value of Σ is:

$$E(\Sigma) = \frac{S_0^{-1}}{df - (r+1)} = \hat{\Sigma}_{OLS}$$

The degrees of freedom are the same as the first prior, that is, $r + 3$. Table 4.7 shows the marginal likelihood estimates and the ALPL obtained using this prior.

Table 4.7: Marginal Likelihood and ALPL for $T - T_0 = 24$, $E(\Sigma) = \hat{\Sigma}_{OLS}$

	Marginal Likelihood				ALPL			
Model	**US**	**UK**	**Japan**	**Brazil**	**US**	**UK**	**Japan**	**Brazil**
M_1	−9499,86	−10 892,89	−11 771,41	−7139,51	−57,93	−59,70	−44,32	−70,60
M_2	−8969,69	−10 322,77	−11 446,59	−6983,87	−42,95	−47,69	−43,06	−67,82
M_3	−8950,27	−10 301,28*	−11 440,66*	−6971,64**	−42,44*	−48,01	−42,90*	−67,49**
M_4	−8899,89*	−10 324,70	−11 441,11	−6988,97	−42,88	−47,43**	−43,23	−68,00
M_5	−8883,16**	−10 299,36**	−11 446,32	−6977,68*	−42,57	−47,55*	−43,10	−67,80
M_6	−8926,9	−10 318,6	−11 436,9**	−6980,8	−41,47**	−47,67	−42,69**	−67,77*

Notes: The best model is marked by ** and the second best by *.

Similar to Table 4.5 the models that account for t-innovations tend to have better performance than those that do not. Regarding the ALPL comparison method, the BVAR-CSV-IG model is best for the US and Japan application, while it is second best for Brazil compared to all models. The moving average specification models perform better than all models for the UK application, where the BVAR-CSV-MA is best followed by BVAR-CSV-MA-t. The BVAR-CSV-t is best for the Brazil application. Table 4.8 shows a similar analysis with an increased cut-off point of 50 periods.

Table 4.8: Marginal Likelihood and ALPL for $T - T_0 = 50$, $E(\Sigma) = \hat{\Sigma}_{OLS}$

	Marginal Likelihood				ALPL			
Model	**US**	**UK**	**Japan**	**Brazil**	**US**	**UK**	**Japan**	**Brazil**
M_1	−9503,30	−10 905,57	−11 767,64	−7138,90	−44,99	−50,80	−42,61	−65,80
M_2	−8977,68	−10 346,62	−11 437,87	−7004,30	−37,58	−43,82*	−41,61	−64,55
M_3	−8960,91	−10 324,74*	−11 439,84	−6986,23**	−37,34	−43,95	−41,60*	−64,27
M_4	−8911,51*	−10 353,05	−11 435,72	−6999,26	−37,48	−43,93	−41,60*	−64,20**
M_5	−8887,29**	−10 322,66**	−11 432,12**	−6989,29*	−37,11*	−43,88	−41,74	−64,31
M_6	−8937,0	−10 339,4	−11 434,3*	−6990,8	−36,91**	−43,75**	−41,51**	−64,20**

Notes: The best model is marked by ** and the second best by *.

Comparing the ALPL for the CSV models over a longer evaluation period, the BVAR-CSV-IG has a better fit for all countries. The BVAR-CSV-MA model performs equally good for the Brazil data.

4.5.3 Financial Data application

We use 1000 observations of daily exchange rate data for 8 currencies that constitute the top trading partners for Zimbabwe in terms of both exports and imports whose ISO Codes are (GBP, EUR, CNY, HKD, INR, ZAR, SGD, ZWD) to the USD. The data for the first 7 currencies was obtained from the Board of Governors of the Federal Reserve and covers the period beginning 29 April 2019 and ending 28 April 2023, while the ZWD series was obtained from the Reserve Bank of Zimbabwe for the same period. Table 4.9 shows the values of the Marginal likelihood for the BVAR-CSV-IG Model and the related standard errors. The evaluation period is 200 for all the exchange rate exercises. As before, the marginal likelihoods have a standard error which is smaller than 1 in all cases.

Table 4.9: BVAR-CSV-IG Model Marginal Likelihoods and Standard Errors Exc

Prior	Marginal Likelihoods(Standard Errors)	
	EXC4	**EXC8**
$E(\Sigma) = I$	15312.5	33953.7
	(0.1158)	(0.0882)
$E(\Sigma) = \hat{\Sigma}_{OLS}$	15437.6	34490.0
	(0.1470)	(0.0855)

Table 4.10 shows the results for a small VAR which estimates the marginal likelihood and ALPL for 4 of the 8 currencies, that is (CNY, ZAR, SGD, ZWD) and a large VAR with all 8 currencies using the prior centred on the Identity matrix.

Table 4.10: Marginal Likelihoods and ALPL $T_0 = 800$, $E(\Sigma) = I$

	Marginal Likelihoods		ALPL	
Model	**EXC4**	**EXC8**	**EXC4**	**EXC8**
BVAR	7461.55	15536.84	9.11	18.89
BVAR-CSV	11332.75	23944.97	10.82*	21.53*
BVAR-CSV-t	12621.33*	27054.99*	N/A	N/A
BVAR-CSV-MA	9770.38	23806.39	N/A	N/A
BVAR-CSV-MA-t	12448.37	24843.39	N/A	N/A
BVAR-CSV-IG	15312.5**	33953.7**	14.42**	33.17**

Notes: The best model is marked by ** and the second best by *. Numbers marked with N/A are numerically very unstable.

The results in the table marked with n.a are numerically very unstable. We obtained very different results in different runs of 150000 iterations each with a burn in of 20000. The BVAR-CSV-IG model has a better fit compared to all models in terms of both the marginal likelihoods and the average log predictive likelihood. This is true for both the smaller VAR with 4 exchange rates and the larger VAR.

Using the alternative prior the results are as follows:

Table 4.11: Marginal Likelihoods and ALPL $T_0 = 800, E(\Sigma) = \hat{\Sigma}_{OLS}$

	Marginal Likelihoods		ALPL	
Model	**EXC4**	**EXC8**	**EXC4**	**EXC8**
BVAR	14 002,95	32 989,11	13,55	31,81
BVAR-CSV	15 335,40	34 408,31	14,35	33,12
BVAR-CSV-t	15 474,51**	34 546,71**	14,47**	33,25**
BVAR-CSV-MA	15 339,91	34 427,03	14,41	33,18
BVAR-CSV-MA-t	15 454,32*	34 474,12	14,42	33,23*
BVAR-CSV-IG	15 437,6	34 490,0*	14,44*	33,20

Notes: The best model is marked by ** and the second best by *.

The BVAR-CSV-t model is best in terms of ALPL for both the small VAR and the large VAR. Comparing the BVAR-CSV model and the BVAR-CSV-IG model alone, given that they have the same number of parameters and the only difference is the distribution we see that the proposed model performs much better in both cases. This we find as compelling evidence for including the property observed in macroeconomic and financial data of heavier tailed distributions.

4.6 Conclusion

We obtained an analytical expression for the likelihood in an inverse gamma CSV model that allows us to obtain a better approximation of the marginal likelihood. Our approach is a result of marginalising out the volatilities and the latent common factor in the model. We compare our approach to popular approaches in the literature using marginal likelihoods and ALPL.

Using 4 macroeconomic applications of datasets for the US, UK, Japan and Brazil that have 20 variables each, we showed that the empirical fit of the common factor inverse gamma SV model is best compared to alternative CSV models in 13 of the 16 data applications. The BVAR-CSV model is best for the Japan data over a longer evaluation period. Over a shorter period using the alternative prior, the BVAR-CSV-MA-t is best for the UK data while the BVAR-CSV-t model is best for Brazil. In the financial application,

the proposed model performs better than alternative models when using the first prior, while the BVAR-CSV-t is best using the alternative prior.

The CSV models with heavier tailed distributions, which is a key property of macroeconomic and financial data, performed much better overall compared to their alternatives thereby indicating strong empirical evidence for using these distributions. The Inverse Gamma CSV model can also be extended to include additional factors, that is to allow the variance Σ_t to vary with time, by using a factor model framework such as proposed in Kim et al. (1998), however, we leave the derivation of the exact likelihood for future research.

CHAPTER 5

R PACKAGE AND TUTORIAL

```
#include <RcppArmadillo.h>
// [[Rcpp::depends(RcppArmadillo)]]
#define ARMA_DONT_PRINT_ERRORS
using namespace Rcpp;
// #define ARMA_NO_DEBUG
#include <omp.h>
// [[Rcpp::plugins(openmp)]]
// [[Rcpp::plugins("cpp11")]]
#include <chrono>
#include <iostream>
#include <random>
#include <cmath>
#include <stdio.h>


//Function to calculate the log of the rising factorial up to p
double lrfact(double n, int p)
{double calcu=std::log(n);  // the result for p=1
  if (p>1){
    for (int ii=1; ii<=(p-1); ii++)
    {calcu=calcu+std::log(n+ii);}
    }
  else if (p==0){
    calcu=std::log(1);
    }
  return calcu;
}

void CalcuLogfac(int niter, int NIT, double n,
                 arma::mat &alogfac, arma::mat &alogfac2,
                 arma::mat &alfac, int nproc)
{ int donde=(niter>NIT)*niter+(NIT>=niter)*NIT;
  omp_set_num_threads(nproc);
  #pragma omp parallel for
 for (int h=0; h<=donde; h++)
  {for (int hold=0; hold<=NIT; hold++)
     {alogfac(hold,h)=lrfact((n+1)*0.5+hold,h);}
     alogfac2(h,0)=lrfact(n*0.5,h);
     alfac(h,0)=lrfact(1,h);
  }
```

```
}

//  Hypergeometric function that uses the rising factorials as inputs
double ourgeoef(int h, arma::mat alogfac, arma::mat alogfac2, arma::mat alfac,
                double zstar, int niter=500)
{ double aux1, aux2, aux3, aux4, aux5, termo;
  aux4=std::log(zstar);
  double sum=1;
  for (int s=1; s<niter; s++)
 {aux1=alogfac(h,s);
  aux2=alogfac(0,s);
  aux3=alogfac2(s,0);
  aux5=alfac(s,0);
  termo=aux1+aux2-aux3+aux4-aux5;
  sum=sum+std::exp(termo);
  aux4=aux4+std::log(zstar);}
return(sum);
}

//Computes the 2F1 Hypergeometric Function
// [[Rcpp::export]]
double ourgeo(double a1, double a2,double b1,double zstar, int niter=500)
{ int s1, s2; double aux1, aux2, aux3, aux4, aux5, termo;
s1=-1+2*(a1>0);
aux1=std::log(std::abs(a1));
aux2=std::log(std::abs(a2));
s2=-1+2*(a2>0);
aux3=std::log(b1);
aux4=std::log(zstar);
aux5=std::log(1);
double sum=1 ;
for (int s=1; s< niter; s++)
{ termo=aux1+aux2-aux3+aux4-aux5;
  sum=sum+s1*s2*std::exp(termo);
  aux1=aux1+std::log(std::abs(a1+s));
  s1=s1*(-1+((a1+s)>0)*2);
  aux2=aux2+std::log(std::abs(a2+s));
  s2=s2*(-1+((a2+s)>0)*2);
  aux3=aux3+std::log((b1+s));
  aux4=aux4+std::log(zstar);
  aux5=aux5+std::log(s+1);
}
return(sum);
}

//Computes the log likelihood for an Inverse Gamma Stochastic Volatility Model
// [[Rcpp::export]]
Rcpp::List lik_clo(arma::mat Res, double b2, int T, double n, double rho, int NIT=300,
                   int niter=300, int nproc=4, int nproc2=4)
```

```
90    //  NIT is the degree of approximation
91  {
92    arma::mat logLik=arma::zeros(T,1);
93    arma::vec oldctil=arma::zeros(NIT+1,1);
94    arma::vec newctil=arma::zeros(NIT+1,1);
95    arma::mat alln=arma::zeros(NIT+1,1);
96    arma::mat allctil=arma::zeros(T,NIT+1);
97    arma::mat allc=arma::zeros(T,NIT+1);
98    int accel=0;  // whether to accelerate or not
99    int donde=(niter>NIT)*niter+(NIT>=niter)*NIT;
100   arma::mat alogfac=arma::zeros(NIT+1,donde+1);
101   arma::mat alogfac2=arma::zeros(donde+1,1);
102   arma::mat alfac=arma::zeros(donde+1,1);
103   omp_set_num_threads(nproc);
104   #pragma omp parallel for
105   for (int h=0; h<=donde; h++)
106   {for (int hold=0; hold<=NIT; hold++)
107   {alogfac(hold,h)=lrfact((n+1)*0.5+hold,h);
108   }
109   alogfac2(h,0)=lrfact(n*0.5,h);
110   alfac(h,0)=lrfact(1,h);
111   }
112   double Vinv=1-rho*rho;
113   double St, Vinvtil, deltaht, normsum, liksum;
114
115   // for t=0
116   double et=Res(0);
117   double useme=b2*et*et;
118   double l0 =-0.5*(n+1)*std::log(0.5*(Vinv+useme));
119   l0 = l0 + 0.5*n*std::log(0.5*Vinv) + (std::lgamma(0.5*(n+1))-std::lgamma(0.5*n))
120      - 0.5*std::log(2*arma::datum::pi)+0.5*log(b2);
121   logLik(0)=l0;
122
123   // for t=1
124   Vinvtil=Vinv+b2*et*et;
125   double bc =(0.5*rho*rho)/(rho*rho+Vinvtil);
126   Vinv=1-rho*rho/(Vinvtil+rho*rho);
127   et=Res(1);
128   double deltah2=rho*rho*1.0/(rho*rho+Vinvtil);
129   double  z=deltah2/(b2*et*et+1);
130   // double  wz=z/(z-1);
131   double  ccc=std::lgamma(0.5*(n+1));
132   double c2= (std::pow(1-z,-0.5*(n+2)))*ourgeo(-0.5,-0.5,0.5*n,z,niter);
133   double c2h=std::log(c2);
134   l0 = -0.5*std::log(2*arma::datum::pi)+0.5*std::log(b2)+ 0.5*(n+1)*std::log(2);
135   l0 = l0 + ccc - 0.5*(n+1)*std::log(b2*et*et+1);
136   normsum = (std::pow(2,0.5*n))*(std::tgamma(0.5*n))*(std::pow(1-deltah2,-0.5*(n+1)));
137   logLik(1)=l0+c2h-log(normsum);
138
```

```
139   // for t=2
140   double S3=1/(b2*et*et+1+rho*rho);
141   double deltah3=(1/(1-S3*rho*rho))*S3*((rho*rho)/(Vinvtil+rho*rho));
142   double c3=(std::pow(1-deltah3,-0.5*(n+2)))*ourgeo(-0.5,-0.5,0.5*n,deltah3,niter)
143    * std::tgamma(0.5*(n+1))*std::pow(1-S3*rho*rho,-0.5*(n+1))*std::pow(2*S3,0.5*(n+1));
144   et=Res(2);
145   Vinvtil=1+b2*et*et;
146   double z3=(S3*rho*rho)/Vinvtil;
147   l0 = -0.5*log(2*arma::datum::pi)+0.5*log(b2)-log(c3);
148   liksum=0;
149   double useful=1/std::pow(Vinvtil,0.5*(n+1))*(std::tgamma(0.5*(n+1))
150   useful=useful/std::tgamma(0.5*(n)))*(std::pow(2,0.5*(n+1))/std::pow(2,0.5*n));
151   useful=std::log(useful);
152   for (int h=0; h<=NIT; h++){
153     double a,b,c, auxx, chat;
154     if (accel)
155     { a=-0.5*(1+2*h); b=-0.5; c=0.5*n;
156     auxx=ourgeo(a,b,c,z3,niter);
157     chat=-0.5*(n+2+2*h)*std::log(1-z3)+std::log(std::abs(auxx));
158     }
159     else {
160       auxx=ourgeoef(h, alogfac,alogfac2, alfac, z3,niter);
161       chat=std::log(std::abs(auxx));
162     }
163     double  c2til=alogfac(0,h)-alogfac2(h,0)+h*std::log(bc)-alfac(h,0);
164     int sign0=-1+2*(auxx>0);
165     double aux1= c2til+ccc+alogfac(0,h)+0.5*(n+1+2*h)*std::log(2*S3)+chat+useful;
166     liksum=liksum + sign0*std::exp(aux1);
167     oldctil(h)=c2til;  // This is the log
168   }
169   logLik(2)=l0+std::log(liksum);
170   allctil.row(1)=arma::trans(oldctil);
171
172   //for all t
173   arma::mat AllVinv=arma::zeros(T,1);   arma::mat Allzt=arma::zeros(T,1);
174   arma::mat AllSt=arma::zeros(T+1,1); arma::mat Alldelta=arma::zeros(T,1);
175   St=S3;
176   AllSt(2)=St;
177   for (int tt=3; tt<=T-1; tt++){
178     double zt;
179     St=1/(b2*et*et+1+rho*rho);
180     deltaht=(1/(1-St*rho*rho))*St*((rho*rho)/(Vinvtil+rho*rho));
181     et=Res(tt);
182     Vinvtil=1+b2*et*et;
183     zt=(St*rho*rho)/Vinvtil;
184     AllSt(tt)=St; Alldelta(tt)=deltaht; AllVinv(tt)=Vinvtil; Allzt(tt)=zt;
185   }
186   arma::mat AllGeo=arma::zeros(T,(NIT+1));
187   #pragma omp parallel for
```

```
188   for (int tt=3; tt<=T-1; tt++){
189     double zt=Allzt(tt);
190     for  (int h=0; h<=NIT; h++){
191       AllGeo(tt,h)=std::log(ourgeoef(h, alogfac,alogfac2, alfac, zt,niter));
192     }
193   }
194   double pSt;
195   l0 = -0.5*std::log(2*arma::datum::pi)+0.5*std::log(b2);
196   for (int tt=3; tt<=T; tt++){
197     double ct,zt;
198     pSt=AllSt(tt-1);
199     if (tt<T){
200       St=AllSt(tt); deltaht=Alldelta(tt); Vinvtil=AllVinv(tt); zt=Allzt(tt);
201     }
202     liksum=0;
203     ct=0;
204     arma::mat allik=arma::zeros(NIT+1,1);
205     arma::mat alct=arma::zeros(NIT+1,1);
206     double NITper=floor((NIT+1)/nproc2);  // iterations per processor
207     int remain=NIT+1-NITper*nproc2;
208     arma::vec nitvec=arma::ones<arma::vec>(nproc2)*NITper;
209     if (remain>0){nitvec.row(nproc2-1)+=remain;}
210     arma::vec limits=arma::cumsum(nitvec);
211     arma::vec trick=arma::zeros(1,1);
212     limits=arma::join_cols(trick, limits);
213     omp_set_num_threads(nproc2);
214     #pragma omp parallel for
215     for (int ii=0; ii<=nproc2-1; ii++){
216       for (int h=limits(ii); h<=(limits(ii+1)-1); h++)
217         {
218         double auxt, chat, ctil, auxi1, auxi2, auxi3;
219         double scale0=alogfac(NIT,h);
220         for (int hold=0; hold<=NIT; hold++)
221         {
222          ctil= oldctil(hold)+ccc+alogfac(0,hold)
223              + alogfac(hold,h)+0.5*(n+1+2*hold)*log(2*pSt)-scale0;
224           newctil(h)=newctil(h)+std::exp(ctil);
225         }
226         newctil(h)= std::log(newctil(h))-alogfac2(h,0)+h*std::log(0.5*rho*rho*pSt)
227                   - alfac(h)+scale0;
228         if (tt<T){
229           auxi1=newctil(h);
230           auxi2=-0.5*(n+1+2*h)*std::log(1-rho*rho*St);
231           auxi3=ccc+alogfac(0,h)+0.5*(n+1+2*h)*std::log(2*St);
232           alct(h)=auxi1+auxi2+auxi3;
233         if (accel){
234           chat= - 0.5*(n+2+2*h)*std::log(1-zt)
235                + std::log(ourgeo(-0.5*(1+2*h),-0.5,0.5*n,zt,niter));}
236           else {
```

```
            chat=AllGeo(tt,h);
          }
          auxt=newctil(h)+ccc+alogfac(0,h)+0.5*(n+1+2*h)*std::log(2*St)+chat;
          allik(h)=auxt;
        }
      }
    }
  if (tt<T){
  double cc0=alct.max()+7;    double ll0=allik.max()+7;
  ct=arma::accu(arma::exp(alct-cc0));
  liksum=arma::accu(arma::exp(allik-ll0));
  liksum=liksum/(std::pow(Vinvtil,0.5*(n+1)))*((std::tgamma(0.5*(n+1))
        /std::tgamma(0.5*(n)))*((std::pow(2,0.5*(n+1)))/std::pow(2,0.5*n)));
  logLik(tt)=l0+std::log(liksum)-std::log(ct)+ll0-cc0;
    }
    newctil=newctil-(newctil(0));
    allctil.row(tt-1)=arma::trans(newctil);
    oldctil=newctil ;
    newctil=arma::zeros(NIT+1,1);
  }

  // For periods 2 and (T+1)
  et=Res(0);
  AllSt(1)=1/(b2*et*et+1);
  et=Res(T-1);
  AllSt(T)=1/(b2*et*et+1);
  double finalLK=arma::accu(logLik);
  double q2=1.0/arma::trace(arma::var(Res));
  double VarLik=-0.5*arma::trace(q2*arma::trans(Res)*Res);
  VarLik+=+0.5*T*std::log(q2);
  VarLik+=-T*0.5*std::log(2*arma::datum::pi);
  return Rcpp::List::create(finalLK, logLik, AllSt, allctil, alogfac,
                            alogfac2, alfac);
}

// Computes the smoothed estimates of the volatility
// [[Rcpp::export]]
arma::vec DrawK0(arma::mat AllSt, arma::mat allctil, arma::mat alogfac,
                 arma::mat alogfac2, arma::mat alfac, int T, int NIT,
                 double n, double rho, double b2,  int nproc2=1)
{ arma::mat allc=arma::zeros(T, NIT+1);
  arma::mat AllW=arma::zeros(T,NIT+1);
  arma::mat AllK=arma::zeros(T,1);
  arma::mat  rowW=arma::zeros(1,NIT+1);
  arma::mat sand;
  // For t=T
  int cualt=T; double sss;
  for (int h=0; h<=(NIT); h++){
    allc(cualt-1,h)=allctil(cualt-1,h);
```

```
286       sss=allc(cualt-1,h)+alogfac(0,h)+0.5*(n+1+2*h)*std::log(2*AllSt(cualt));
287       rowW(0,h)=sss;
288     }
289     double sss0=rowW.max()+7;
290     rowW=arma::exp(rowW-sss0);
291     sss=arma::accu(rowW);
292     rowW=rowW*(1.0/sss);
293     AllW.row(cualt-1)=rowW;
294     std::mt19937 generator((std::chrono::steady_clock::now().time_since_epoch().count()));
295     std::uniform_real_distribution<double> distu(0.0,1.0);
296     double drawu, Khere, fredom; int where;
297     drawu=distu(generator);
298     sand=arma::cumsum(arma::trans(rowW));
299     where=std::round(arma::accu(sand<drawu));  // This gives indices from 0 to NIT+1
300     fredom=(n+1)*0.5;
301     std::gamma_distribution<double> distgam((fredom+where),2*AllSt(cualt));
302     Khere=distgam(generator);
303     AllK(cualt-1)=Khere;
304     cualt=T-1; // for t=T-1, ..., 1
305     omp_set_num_threads(nproc2);
306     for (int cualt=(T-1); cualt>=1; cualt--)
307     {
308   #pragma omp parallel for
309       for (int h=0; h<=(NIT); h++){
310         double auxil1=0; double auxil2, sss1;
311         if (cualt>1){
312           for (int ht=0; ht<=h; ht++)
313           {
314             auxil2 = allctil(cualt-1,h-ht)+ht*std::log(Khere)+ht*std::log(0.25*rho*rho)
315                    - alogfac2(ht,0)-alfac(ht,0);
316             auxil1+=std::exp(auxil2);
317           }
318         }
319         else if (cualt==1)
320         {
321           auxil1=h*std::log(Khere)+h*std::log(0.25*rho*rho)-alogfac2(h,0)-alfac(h,0);
322           auxil1=std::exp(auxil1);
323         }
324         allc(cualt-1,h)=std::log(auxil1);
325         sss1=allc(cualt-1,h)+alogfac(0,h)+0.5*(n+1+2*h)*std::log(2*AllSt(cualt));
326         rowW(0,h)=sss1;
327       }
328       sss0=rowW.max()+7;
329       rowW=arma::exp(rowW-sss0);
330       sss=arma::accu(rowW);
331       rowW=rowW*(1.0/sss);
332       AllW.row(cualt-1)=rowW;
333       drawu=distu(generator);
334       sand=arma::cumsum(arma::trans(rowW));
```

```
335       where=std::round(arma::accu(sand<drawu));  //indices from 0 to NpartHere-1
336       std::gamma_distribution<double> distgam((fredom+where),2*AllSt(cualt));
337       Khere=distgam(generator);
338       AllK(cualt-1)=Khere;
339     }
340     return AllK;
341 }
342
343 /*** R
344 */
345
346
```

5.1 Package description and installation

The main function **lik_clo** computes the log likelihood for an inverse gamma stochastic volatility model using a closed form expression of the likelihood. The closed form expression is obtained for the log likelihood of a stationary inverse gamma stochastic volatility model by marginalising out the volatilities. This allows the user to obtain the maximum likelihood estimator for this non linear non Gaussian state space models. In addition, we can obtain the smoothed estimates of the volatility using draws from the exact posterior distribution of the inverse volatility by invoking the function **DrawK0**. Lastly one can evaluate the 2F1 hypergeometric function using **ourgeo**. The package can be installed in R as:

```
install.packages("invgamstochvol")
```

and using the library

```
library(invgamstochvol)
```

5.1.1 lik_clo

The function computes the log likelihood for an inverse gamma stochastic volatility model using a closed form expression of the likelihood.

Usage

lik_clo(Res, b2, n, rho, NIT=200, niter=200, nproc=2, nproc2=2)

Arguments

Res	Matrix of OLS residuals, usually resulting from a call to priorvar.
b2	Level of volatility.
n	Degrees of freedom
rho	The parameter for the persistence of volatility.
NIT	The degree of approximation to truncate the log likelihood sum. The default value is set at 200.
niter	The degree of approximation to truncate the hypergeometric sum. The default value is set at 200.
nproc	The number of processors allocated to evaluating the hypergeometric function. The default value is set at 2. Increase this value to speed up the processes.
nproc2	The number of processors allocated to computing the log likelihood. The default value is set at 2. Increase this value to speed up the processes.

The function returns a list of 7 items.List item number 1, is the sum of the log likelihood, while the rest are constants that are useful to obtain the smoothed estimates of the volatility.

Example

We provide example data used in this paper with the package. This data is for the US inflation data. The data frame has 247 observations for the period 1960Q1 to 2022Q3. The following example obtains the likelihood for this data:

```
##Example using US data
data1 <- US_Inf_Data
Ydep <- as.matrix(data1)
littlerho=0.95
r0=1
rho=diag(r0)*littlerho
p=4
n=4.1
T=nrow(Ydep)
Xdep <- Ydep[p:(T-1),]
if (p>1){
  for(lagi in 2:p){
    Xdep <- cbind(Xdep, Ydep[(p-lagi+1):(T-lagi),])
  }
```

```
}
T=nrow(Ydep)
Ydep <- as.matrix(Ydep[(p+1):T,])
T=nrow(Ydep)
unos <- rep(1,T)
Xdep <- cbind(unos, Xdep)

## obtain residuals
bOLS <- solve(t(Xdep) %*% Xdep) %*% t(Xdep) %*% Ydep
Res= Ydep- Xdep %*% bOLS
Res=Res[1:T,1]
b2=solve(t(Res) %*% Res/T) %*% (1-rho %*% rho)/(n-2)
Res=as.matrix(Res,ncol=1)

##obtain the log likelihood
LL1=lik_clo(Res,b2,n,rho)
```

Maximum likelihood estimation can then be used to obtain the value of the likelihood at the maximum and parameter estimates.

5.1.2 DrawK0

The function obtains the smoothed estimates of the volatility, by obtaining a random draw from the exact posterior of the inverse volatilities.

Usage

DrawK0(AllSt, allctil, alogfac, alogfac2, alfac, n, rho, b2, nproc2=2)

Arguments

AllSt	Some constants obtained from the evaluation of the log likelihood using the function lik_clo.
allctil	Some constants obtained from the evaluation of the log likelihood using the function lik_clo.
alogfac	Some constants obtained from the evaluation of the log likelihood using the function lik_clo.
alogfac2	Some constants obtained from the evaluation of the log likelihood using the function lik_clo.
alfac	Some constants obtained from the evaluation of the log likelihood using the function lik_clo.

n	Degrees of freedom.
rho	The parameter for the persistence of volatility.
b2	Level of volatility.
nproc2	The number of processors allocated to the calculations. The default value is set at 2.

The function returns a vector with a random draw from the posterior of the inverse volatilities. To obtain the smoothed estimates of the volatility using this package, one has to save the constants obtained from evaluating the function lik_clo as follows:

Example

```
deg=200
niter=200
AllSt=matrix(unlist(LL1[3]), ncol=1)
allctil=matrix(unlist(LL1[4]),nrow=T, ncol=(deg+1))
donde=(niter>deg)*niter+(deg>=niter)*deg
alogfac=matrix(unlist(LL1[5]),nrow=(deg+1),ncol=(donde+1))
alogfac2=matrix(unlist(LL1[6]), ncol=1)
alfac=matrix(unlist(LL1[7]), ncol=1)
```

Then by averaging draws from the exact posterior distribution of the inverse volatilities, the smoothed estimates of the volatility can be obtained.

```
milaK=0
repli=5
keep0=matrix(0,nrow=repli, ncol=1)
for (jj in 1:repli)
{
  laK=DrawK0(AllSt,allctil,alogfac, alogfac2, alfac, n, rho, b2,nproc2=2)

  milaK=milaK+1/laK*(1/repli)
  keep0[jj]=mean(1/laK)/b2
}
ccc=1/b2
fefo=as.vector(milaK)*ccc
##obtain moving average of squared residuals
mRes=matrix(0,nrow=T,ncol=1)
Res2=Res*Res
bandi=5
for (iter in 1:T)
{  low=(iter-bandi)*(iter>bandi)+1*(iter<=bandi)
   up=(iter+bandi)*(iter<=(T-bandi))+T*(iter>(T-bandi))
  mRes[iter]=mean(Res2[low:up])
```

```
}

##plot the results
plot(fefo,type="l", col = "red", xlab="Time",ylab="Volatility Means")
lines(mRes, type="l", col = "blue")
legend("topright", legend = c("Stochastic Volatility", "Squared Residuals"),
       col = c("red", "blue"), lty = 1, cex = 0.8)

```

where repli is the number of replications. Lastly, to evaluate a 2F1 hypergeometric function, one can use the function **ourgeo** as follows:

5.1.3 ourgeo

The function computes the 2F1 Hypergeometric Function.

Usage

ourgeo(a1, a2, b1, zstar, *niter* = 500)

Arguments

a1	Parameter (Real)
a2	Parameter (Real)
b1	Parameter (Real)
zstar	Primary real argument
niter	The degree of approximation to truncate the hypergeometric sum. The default value is set at 500.

The function returns the value of the hypergeometric function

Example

```
ourgeo(1.5,1.9,1.2,0.7)
```

CHAPTER 6
CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

6.1 Introduction

In this chapter, we present the conclusions, limitations, policy implications for the studies covered in the dissertation. The policy implications are derived from the findings in all the chapters. In addition, this chapter suggests directions for future research.

6.2 Conclusion

We propose a novel approach to obtaining the analytical expressions of the likelihood for stationary inverse gamma Stochastic Volatility models. In the absence of analytical expressions of the log likelihood, it is not possible to obtain the Maximum Likelihood Estimator (MLE) for this class of non linear non gaussian state space models for the univariate model. In addition, the smoothed estimates of the volatility can not be obtained. We use the approach to obtaining the likelihood in this dissertation to analyse a univariate inverse gamma stochastic volatility model and a multivariate approximate factor model resulting in the two studies covered in this dissertation.

First, chapter 3 considers the univariate inverse gamma stochastic volatility model and obtains the exact likelihood for this model. This expression of the likelihood allows the estimation of the parameters and unobserved states for this model class by maximum likelihood. Further, we provide the analytical expressions for both the filtering and smoothing distributions of the volatilities obtained as mixture of gammas and therefore we provide the smoothed estimates of the volatility. The chapter shows that by marginalising out the volatilities, the model that is obtained has the resemblance of a GARCH in the sense that the formulas are similar, which simplifies computations significantly.

To demonstrate this approach and to compare its performance in the univariate case,

we provide two empirical applications using financial and macroeconomic variables. The macroeconomic application uses quarterly inflation data for four countries, that is, UK, USA, Japan and Brazil. A range of other models are also estimated to evaluate the empirical performance of the proposed model. We find that the proposed model performs better than all other models in 50% of the applications in terms of the Bayesian Information Criterion, with very large gains for the Brazil dataset. The second application uses exchange rates data for 7 currencies (GBP, EUR, JPY, CND, AUD, BRL, ZAR) and finds that the empirical fit of the proposed model is overall better than alternative models in 4 of the 7 currencies in terms of the BIC, being much superior in the case of currencies with turbulent episodes, such as the Brazilian Real.

In Chapter 4, we extend the univariate approach to estimating large Vector Autoregressions (VAR) in a multivariate stochastic volatility model. Given the increase in fat tailed events over the years and the empirical evidence on commonality that is observed in volatility patterns, this model combines stochastic volatility, heavy tailed distributions and a common latent factor. The common latent factor is taken as multiplicative. This model is estimated using 4 macroeconomic applications that use 20 variables each for Japan, Brazil, US, and the UK. A second application uses financial data of daily exchange rate returns for a small VAR of 4 currencies and a larger VAR of 8 currencies. The comparison method is based on marginal likelihoods and the one step ahead out of sample predictive likelihoods. The proposed model is compared to other common stochastic volatility models and a factor stochastic volatility model with varying number of factors.

The empirical fit of the common factor inverse gamma SV model fares very significantly better compared to other common stochastic volatility models for both the macro economic and financial application. Using 4 macroeconomic applications of datasets for the US, UK, Japan and Brazil that have 20 variables each, we showed that the empirical fit of the common factor inverse gamma SV model is best compared to alternative CSV models in 13 of the 16 data applications. The BVAR-CSV model is best for the Japan data over a longer evaluation period. Over a shorter period using the alternative prior, the BVAR-CSV-MA-t is best for the UK data while the BVAR-CSV-t model is best for Brazil. In the financial application, the proposed model performs better than alternative

models when using the first prior, while the BVAR-CSV-t is best using the alternative prior.

The CSV models with heavier tailed distributions, which is a key property of macroeconomic and financial data, performed much better overall compared to their alternatives thereby indicating strong empirical evidence for using these distributions.The BVAR-CSV-IG model performed substantially better than all the models. These findings suggest that the BVAR-CSV-IG specification significantly improves computations of models with a common stochastic volatility specification and therefore, further research into these models would be advantageous.

We provide the computer code publicly as an R package on the CRAN repository that is used to obtain the likelihood for a univariate inverse gamma stochastic volatility and also the smoothed estimates of the volatility. The code and tutorial is detailed in chapter 5 above.

6.3 Policy Implications

Given that policy makers rely on forecasts to assess the impact of reforms ex-ante, the findings from the empirical applications in this dissertation are of paramount importance. To start with, central banks across the economies tend to rely on predictive densities to illustrate and project variables such as inflation levels and the GDP growth rate among other variables. The models proposed in the dissertation aim to provide alternative models that improve forecasting accuracy while at the same time remaining simple and efficient. The comparison exercise results in both the univariate and multivariate applications showed that a number of datasets tend to prefer the inverse gamma specification as it improves forecasting accuracy marginally and thus, will improve density forecasts. This inverse gamma specification can account for fat tails that are observed in most macroeconomic and financial data. Thus, central banks and other macroeconomic institutions would largely benefit from using these models and approaches that we suggest in their forecasting exercises. We provide the computer code publicly as an R package on the CRAN repository that is used to obtain the likelihood for a univariate inverse gamma stochastic volatility that relevant institutions can use in their various exercises.

This R package also allows the user to obtain the smoothed estimates of the volatility as an estimate of the variance of e_t at each point of time given all available data that will be useful in forecasting exercises by the relevant authorities.

6.4 Limitations and Directions for Further Research

While the models, derivations and analysis proposed in this dissertation are robust, they are not without limitations. Considering the univariate stochastic volatility model in chapter 3 for instance, one such limitation is that for some datasets characterised with sporadic marginal jumps in volatility such as the Zimbabwean Dollar exchange rate, the analysis will fail to converge to a global maximum at least without further manipulations to the algorithm. However, this limitation is not peculiar to the inverse gamma model alone but all models that were used in the comparison exercise in this chapter.

Regarding the study in chapter 4, the main limitation is a result of the restrictions on hyperparameters for the minnesota priors used in common stochastic volatility models that are fixed at some values mentioned in the discussion of the prior values. These restrictions though seemingly restrictive, are supported nonetheless by empirical evidence on the co-movements observed in the volatility of macroeconomic variables.

In future research, it would be useful to extend the proposed approach to allow Σ to also vary with time. We consider this approach in future empirical exercises. In addition, it may be useful to extend the current approach to allow for the estimation of more complex data such as the data that has jumps in volatility which will especially be useful in forecasting exercises for developing economies.

Appendix

Appendix

A.1 Appendices to Chapter 3

A.1.1 Proof of Lemma

To derive the likelihood we will make use of the following lemma, which is a slightly modified version of Theorem 7.3.4 in Muirhead (2005).

Lemma A.1.1. *For integers $p \leq q$*

$$\int |K|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}AK\right) {}_pF_q\left(a_1, ..., a_p; b_1, ..., b_q; \frac{1}{4}BK\right) dK = \Gamma\left(\frac{n+1}{2}\right) \left|\frac{1}{2}A\right|^{-\frac{n+1}{2}} \times {}_{p+1}F_q\left(a_1, ..., a_p, \frac{n+1}{2}; b_1, ..., b_q; \frac{1}{2}BA^{-1}\right)$$

where $(n+1)/2 > 0$ and ${}_pF_q(.)$ is a hypergeometric function of scalar argument, provided that in the case $p = q$ we have that $|0.5BA^{-1}| < 1$.

Proof. We apply Theorem 7.3.4 in Muirhead (2005) after making a change of variables. Let $X = \frac{1}{4}BK$ such that $K = 4XB^{-1}$. Thus we have:

$${}_pF_q\left(a_1, ..., a_p; b_1, ..., b_q; \frac{1}{4}BK\right) = {}_pF_q\left(a_1, ..., a_p; b_1, ..., b_q; X\right)$$

Therefore the integral becomes as follows:

$$\int |X|^{\frac{n+1-2}{2}} |4B^{-1}|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}4AXB^{-1}\right) {}_pF_q\left(a_1, ..., a_p; b_1, ..., b_q; X\right) dK$$

We use the Jacobian $dK = |4B^{-1}|dX$ to integrate with respect to X:

$$\int |X|^{\frac{n+1-2}{2}} \exp(-2XB^{-1}A) {}_pF_q\left(a_1, ..., a_p; b_1, ..., b_q; X\right) dX |4B^{-1}|^{\frac{n+1}{2}}$$

This integral is the same as in the theorem, therefore, when we integrate out X we get the following:

$$\int \frac{|X|^{\frac{n+1-2}{2}}}{|4B^{-1}|^{-\frac{n+1}{2}}} \exp(-X2B^{-1}A)_pF_q\bigg(a_1,...,a_p;b_1,...,b_q;X\bigg)dX = \Gamma\bigg(\frac{n+1}{2}\bigg)\bigg|\frac{1}{2}A\bigg|^{-\frac{n+1}{2}} \times {}_{p+1}F_q\bigg(a_1,...,a_p,\frac{n+1}{2};b_1,...,b_q;\frac{1}{2}BA^{-1}\bigg)$$

□

A.1.2 Proof of Proposition 3.3.1

Proof. k_1 is a gamma, Bauwens et al. (2000) gives the prior density for k_1 as:

$$|k_1|^{\frac{n-2}{2}} \exp\bigg(-\frac{1}{2}\big(k_1(1-\rho^2)\big)\bigg)\frac{1}{c_0} \tag{A.1.1}$$

where $c_0 = \frac{\Gamma(\frac{n}{2})}{(\frac{1-\rho^2}{2})^{\frac{n}{2}}}$, is a constant and Γ is a gamma function. Let $V_1^{-1} = (1-\rho^2)$, thus, the likelihood for the first observation is as follows:

$$\begin{aligned} L(y_1) &= \int L(y_1 \mid k_1)\pi(k_1)dk_1 \\ &= \int (2\pi)^{-\frac{1}{2}}\sqrt{B^2}k_1^{\frac{1}{2}} \exp\bigg(-\frac{1}{2}e_1^2B^2k_1\bigg)k_1^{\frac{n-2}{2}} \exp\bigg(-\frac{1}{2}(1-\rho^2)k_1\bigg)\frac{1}{c_0}dk_1 \end{aligned} \tag{A.1.2}$$

The integral is with respect to k_1, so after rearranging and combining like terms we have;

$$L(y_1) = \int (2\pi)^{-\frac{1}{2}}\sqrt{B^2}k_1^{\frac{n+1-2}{2}} \exp\bigg(-\frac{1}{2}(B^2e_1^2 + V_1^{-1})k_1\bigg)\frac{1}{c_0}dk_1$$

where $k_1^{\frac{n+1-2}{2}} \exp(-\frac{1}{2}(B^2e_1^2 + V_1^{-1})k_1)$ is the kernel of a gamma with $n+1$ degrees of freedom. Let $\widetilde{V}_2 = (B^2e_1^2 + V_1^{-1})^{-1}$, therefore, the density of $k_1|y_1$ is:

$$\pi(k_1|y_1) = k_1^{\frac{n+1-2}{2}} \exp\bigg(-\frac{1}{2}k_1\widetilde{V}_2^{-1}\bigg)\frac{1}{\bar{c_0}} \tag{A.1.3}$$

with $\bar{c_0} = \frac{\Gamma(\frac{n+1}{2})}{(\frac{\widetilde{V}_2^{-1}}{2})^{\frac{n+1}{2}}}$. Thus, we have the likelihood as:

$$L(y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\Gamma\bigg(\frac{n+1}{2}\bigg)\bigg|\frac{B^2e_1^2 + V_1^{-1}}{2}\bigg|^{-\frac{n+1}{2}}\frac{1}{c_0}$$

Taking into account c_0 we can write the likelihood for $t = 1$ as:

$$L(y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}2^{\frac{1}{2}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\left|B^2e_1^2 + V_1^{-1}\right|^{-\frac{n+1}{2}}V_1^{-\frac{n}{2}}$$

Define $k_{1:2} = (k_1, k_2)$, then we have the likelihood for the second observation as:

$$L(y_2|y_1) = \int L(y_2|k_{1:2}, y_1)\pi(k_{1:2}|y_1)dk_{1:2}$$

where $\pi(k_{1:2}|y_1) = \pi(k_1|y_1)\pi(k_2|k_1, y_1)$. The prior for k_t unconditionally is a gamma. However, $k_t|k_{t-1}$ is a non central chi-squared. Muirhead (2005, p. 442) gives this non central chi-squared density as follows:

$$\pi(k_t|k_{t-1}) = k_t^{\frac{n-2}{2}}\exp\left(-\frac{1}{2}k_t\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2k_{t-1}k_t\right)\exp\left(-\frac{1}{2}\rho^2k_{t-1}\right)\left(\Gamma\left(\frac{n}{2}\right)\right)^{-1}\frac{1}{c} \tag{A.1.4}$$

where ${}_0F_1$ is a hypergeometric function, ρ^2k_{t-1} is the non-centrality parameter and $c = 2^{\frac{n}{2}}$. We can then write the likelihood for the second observation given the first as :

$$L(y_2|y_1) = \int(2\pi)^{-\frac{1}{2}}\sqrt{B^2}k_2^{\frac{1}{2}}\exp\left(-\frac{1}{2}B^2e_2^2k_2\right)\pi(k_{1:2}|y_1)dk_{1:2} \tag{A.1.5}$$

We integrate first with respect to k_1. Define l_2 as representing all the elements in $\pi(k_2|k_1)$ as given by (A.1.4) that do not depend on k_1 as follows:

$$l_2 = \left(k_2^{\frac{n-2}{2}}\exp\left(-\frac{1}{2}k_2\right)\right)^{-1}\left(\frac{1}{\Gamma(\frac{n}{2})}\right)^{-1}\left(\frac{1}{c}\right)^{-1} \tag{A.1.6}$$

Given that $\pi(k_2|k_1, y_1) = \pi(k_2|k_1)$, and given (A.1.4) and (A.1.3), we can write $\pi(k_2|y_1)$ as follows:

$$\pi(k_2|y_1) = \int\pi(k_2|k_1, y_1)\pi(k_1|y_1)dk_1 =$$
$$\int\frac{1}{\bar{c_0}}k_1^{\frac{n+1-2}{2}}\exp\left(-\frac{1}{2}(\widetilde{V_2}^{-1}k_1)\right)\exp\left(-\frac{1}{2}(\rho^2k_1)\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2k_1k_2\right)\frac{1}{l_2}dk_1$$

where we have used the expression for $\pi(k_1|y_1)$ in (A.1.3). We can write the above integral more compactly as:

$$\int \pi(k_2|k_1,y_1)\pi(k_1|y_1)dk_1 = \int \frac{1}{\bar{c_0}} k_1^{\frac{n+1-2}{2}} \exp\bigg(-\frac{1}{2}(\widetilde{V_2}^{-1}+\rho^2)k_1\bigg)_0F_1\bigg(\frac{n}{2};\frac{1}{4}\rho^2k_1k_2\bigg)\frac{1}{l_2}dk_1$$

Applying Lemma 6.1 the solution to this integral is as follows:

$$\begin{aligned}\pi(k_2|y_1) = &\int \pi(k_2|k_1,y_1)\pi(k_1|y_1)dk_1 = \\ &\frac{1}{\bar{c_0}}\Gamma\bigg(\frac{n+1}{2}\bigg)\bigg|\frac{\widetilde{V_2}^{-1}+\rho^2}{2}\bigg|^{-\frac{n+1}{2}} {}_1F_1\bigg(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\bigg)\frac{1}{l_2}\end{aligned} \tag{A.1.7}$$

Given (A.1.6) and (A.1.7), the distribution of $k_2|y_1$ is a mixture of gammas as follows:

$$\pi(k_2|y_1) \propto k_2^{\frac{n-2}{2}} \exp\bigg(-\frac{1}{2}k_2\bigg)_1F_1\bigg(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\bigg) \tag{A.1.8}$$

The normalising constant for this density function can be obtained in closed form. Lemma 6.1 gives the solution to this integral, thus, we have:

$$\int k_2^{\frac{n-2}{2}} \exp\bigg(-\frac{1}{2}k_2\bigg)_1F_1\bigg(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}k_2\delta_2\bigg)dk_2 = \Gamma\bigg(\frac{n}{2}\bigg)2^{\frac{n}{2}}{}_2F_1\bigg(\frac{n+1}{2},\frac{n}{2};\frac{n}{2};\delta_2\bigg) \tag{A.1.9}$$

where $\delta_2 = \rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}$. This ${}_2F_1\big(\frac{n+1}{2},\frac{n}{2};\frac{n}{2};\delta_2\big)$ function has the same terms in the denominator and the numerator thus they cancel out and we have:

$${}_2F_1\bigg(\frac{n+1}{2},\frac{n}{2};\frac{n}{2};\delta_2\bigg) = {}_1F_0\bigg(\frac{n+1}{2};\delta_2\bigg) \tag{A.1.10}$$

Therefore, this function simplifies to a known solution for $|\delta_2| < 1$, see Muirhead (2005, p.261) .

$${}_1F_0\bigg(\frac{n+1}{2};\delta_2\bigg) = (1-\delta_2)^{-\frac{n+1}{2}} \tag{A.1.11}$$

Therefore the normalising constant becomes:

$$\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}{}_1F_0\left(\frac{n+1}{2};\delta_2\right)=\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}(1-\delta_2)^{-\frac{n+1}{2}}$$

Given this normalising constant, we have the density for $\pi(k_2|y_1)$ from A.1.8 as follows:

$$\pi(k_2|y_1)=\frac{1}{c_1}k_2^{\frac{n-2}{2}}\exp\left(-\frac{1}{2}k_2\right){}_1F_1\left(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\right)$$

where $c_1=\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}(1-\delta_2)^{-\frac{n+1}{2}}$. Thus, the likelihood for the second observation is as follows:

$$\begin{aligned}L(y_2|y_1)&=\int\pi(y_2|k_2,y_1)\pi(k_2|y_1)dk_2\\&=\int(2\pi)^{-\frac{1}{2}}\sqrt{B^2}k_2^{\frac{n+1-2}{2}}\exp\left(-\frac{1}{2}(B^2e_2^2+1)k_2\right)\frac{1}{c_1}{}_1F_1\left(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\right)dk_2\end{aligned}$$

Using again Lemma 6.1 and taking into account c_1, the likelihood for the second observation is:

$$L(y_2|y_1)=(2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{(B^2e_2^2+1)^{-\frac{n+1}{2}}}{(1-\delta_2)^{-\frac{n+1}{2}}}{}_2F_1\left(\frac{n+1}{2},\frac{n+1}{2};\frac{n}{2};(B^2e_2^2+1)^{-1}\delta_2\right)$$

Thus we get a Gauss hypergeometric function which can be evaluated easily. Let $Z_2=(B^2e_2^2+1)^{-1}\delta_2$ and $\hat{C}_2={}_2F_1\left(\frac{n+1}{2},\frac{n+1}{2};\frac{n}{2};Z_2\right)$. This series converges because $|Z_2|<1$ (Abramowitz et al., 1988). To accelerate the convergence of this series we apply the Euler transformation as in Abramowitz et al. (1988) and thus we get:

$${}_2F_1\left(\frac{n+1}{2},\frac{n+1}{2};\frac{n}{2};Z_2\right)=(1-Z_2)^{-\frac{n+2}{2}}{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{n}{2};Z_2\right)\qquad\text{(A.1.12)}$$

Thus $\hat{C}_2={}_2F_1\left(\frac{n+1}{2},\frac{n+1}{2};\frac{n}{2};Z_2\right)=(1-Z_2)^{-\frac{n+2}{2}}{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{n}{2};Z_2\right)$, then we can write the $L(y_2|y_1)$ as follows:

$$L(y_2|y_1)=(2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{(B^2e_2^2+1)^{-\frac{n+1}{2}}}{(1-\delta_2)^{-\frac{n+1}{2}}}\hat{C}_2$$

The density of k_t for the third observation is given by:

$$\pi(k_3|y_2,y_1)=\int \pi(k_3|k_2)\pi(k_2|y_2,y_1)dk_2$$

where $\pi(k_2|y_2,y_1)\propto\pi(k_2|y_1)L(y_2|k_2,y_1)$. The distribution for $\pi(k_2|y_1)$ in (A.1.8) can be written as follows:

$$\pi(k_2|y_1)\propto\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}k_2^{\frac{n+2h_2-2}{2}}\exp\Big(-\frac{1}{2}k_2\Big)$$

where $\widetilde{C}_{2,h_2}=\frac{[(n+1)/2]_{h_2}}{[n/2]_{h_2}}\big(\frac{1}{2}\rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}\big)^{h_2}\frac{1}{h_2!}$. Thus we have:

$$\pi(k_2|y_2,y_1)\propto\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}k_2^{\frac{n+1+2h_2-2}{2}}\exp\Big(-\frac{1}{2}k_2(B^2e_2^2+1)\Big) \tag{A.1.13}$$

Given (A.1.4) and (A.1.13) we have:

$$\begin{aligned}\pi(k_3|y_2,y_1)\propto\int k_3^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_3\Big){}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2k_2k_3\Big)\exp\Big(-\frac{1}{2}\rho^2k_2\Big)\\ \times\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}k_2^{\frac{n+1+2h_2-2}{2}}\exp\Big(-\frac{1}{2}k_2(B^2e_2^2+1)\Big)\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_2\end{aligned}$$

which simplifies to:

$$\begin{aligned}\pi(k_3|y_2,y_1)\propto\int k_3^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_3\Big){}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2k_2k_3\Big)\exp\Big(-\frac{1}{2}(B^2e_2^2+1+\rho^2)k_2\Big)\\ \sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}k_2^{\frac{n+1+2h_2-2}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_2\end{aligned}$$

Using Lemma 6.1 the density of $k_3|y_2,y_1$ is thus:

$$\begin{aligned}\pi(k_3|y_2,y_1)=\frac{1}{c_3}k_3^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_3\Big)\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\Big(\frac{n+1+2h_2}{2}\Big)\\ {}_1F_1\Big(\frac{n+1+2h_2}{2};\frac{n}{2};\frac{1}{2}k_3\rho^2S_3\Big)(2S_3)^{\frac{n+1+2h_2}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}\end{aligned} \tag{A.1.14}$$

where $S_3=(B^2e_2^2+1+\rho^2)^{-1}$ and c_3 is the normalising constant as in (A.1.9) as follows:

$$c_3 = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+1+2h_2}{2}\right) (2S_3)^{\frac{n+1+2h_2}{2}} {}_2F_1\left(\frac{n+1+2h_2}{2}, \frac{n}{2}; \frac{n}{2}; \rho^2 S_3\right)$$

Similar to (A.1.10) and (A.1.11), the hypergeometric function simplifies to get:

$$c_3 = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+1+2h_2}{2}\right) (2S_3)^{\frac{n+1+2h_2}{2}} (1-\rho^2 S_3)^{-\frac{n+1+2h_2}{2}}$$

Collecting terms dependent on h_2 we can write c_3 as

$$c_3 = \left(\sum_{h_2=0}^{\infty} \frac{[(n+1)/2]_{h_2}}{[n/2]_{h_2}} [(n+1)/2]_{h_2} \frac{\delta_3^{h_2}}{h_2!}\right) \Gamma\left(\frac{n+1}{2}\right) (1-\rho^2 S_3)^{-\frac{n+1}{2}} (2S_3)^{\frac{n+1}{2}}$$

where $\delta_3 = \left((1-\rho^2 S_3)^{-1} S_3 \rho^2 (\widetilde{V_2}^{-1} + \rho^2)^{-1}\right)$. This can be written as:

$$c_3 = {}_2F_1\left(\frac{n+1}{2}, \frac{n+1}{2}; \frac{n}{2}; \delta_3\right) \Gamma\left(\frac{n+1}{2}\right) (1-\rho^2 S_3)^{-\frac{n+1}{2}} (2S_3)^{\frac{n+1}{2}}$$

Using Euler's acceleration in (A.1.12) we can transform c_3 as:

$$c_3 = (1-\delta_3)^{-\frac{n+2}{2}} {}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{n}{2}; \delta_3\right) \Gamma\left(\frac{n+1}{2}\right) (1-\rho^2 S_3)^{-\frac{n+1}{2}} (2S_3)^{\frac{n+1}{2}}$$

Therefore the likelihood for $t = 3$ is as follows:

$$L(y_3|y_2, y_1) = \int \pi(y_3|k_3, y_2, y_1) \pi(k_3|y_2, y_1) dk_3$$

Thus we have from (A.1.14)

$$L(y_3|y_2, y_1) = \int (2\pi)^{-\frac{1}{2}} \sqrt{B^2} \frac{1}{c_3} k_3^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2} k_3 (B^2 e_3^2 + 1)\right) \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+1+2h_2}{2}\right) {}_1F_1\left(\frac{n+1+2h_2}{2}; \frac{n}{2}; \frac{1}{2} k_3 \rho^2 S_3\right) (2S_3)^{\frac{n+1+2h_2}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} dk_3$$

and using Lemma 6.1 we get:

$$L(y_3|y_2,y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+1+2h_2}{2}\right)(2S_3)^{\frac{n+1+2h_2}{2}}\Gamma\left(\frac{n+1}{2}\right)2^{\frac{n+1}{2}}$$
$$(B^2e_3^2+1)^{-\frac{n+1}{2}}{}_2F_1\left(\frac{n+1+2h_2}{2},\frac{n+1}{2};\frac{n}{2};(B^2e_3^2+1)^{-1}\rho^2S_3\right)\frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}$$

Letting $Z_3 = (B^2e_3^2+1)^{-1}\rho^2S_3$, we can define $\hat{C}_3 = {}_2F_1\left(\frac{n+1+2h_2}{2},\frac{n+1}{2};\frac{n}{2};Z_3\right)$. Thus, we have:

$$L(y_3|y_2,y_1) = (2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\frac{\Gamma\left(\frac{n+1+2h_2}{2}\right)}{(B^2e_3^2+1)^{\frac{n+1}{2}}}(2S_3)^{\frac{n+1+2h_2}{2}}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_3$$

The filtering density of k_t for $t=4$ is given by:

$$\pi(k_4|y_3,y_2,y_1) = \int \pi(k_4|k_3,y_1,y_2,y_3)\pi(k_3|y_3,y_2,y_1)dk_3 \tag{A.1.15}$$

where $\pi(k_3|y_3,y_2,y_1) \propto \pi(k_3|y_2,y_1)L(y_3|y_2,y_1)$. Let:

$$\widetilde{C}_{3,h_3} = \sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+1+2h_2}{2}\right)\frac{[(n+1)/2+h_2]_{h_3}}{[n/2]_{h_3}}\left(\frac{1}{2}\rho^2S_3\right)^{h_3}\frac{1}{h_3!}(2S_3)^{\frac{n+1+2h_2}{2}} \tag{A.1.16}$$

Then from (A.1.14) we have that the filtering distribution $k_3|y_2,y_1$ is a mixture of gammas as follows:

$$\pi(k_3|y_2,y_1) \propto \sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}k_3^{\frac{n+2h_3-2}{2}}\exp\left(-\frac{1}{2}k_3\right)$$

As before, when we include the third observation, the distribution of $k_3|y_3,y_2,y_1$ is also a mixture of gammas and can be written as follows:

$$\pi(k_3|y_3,y_2,y_1) \propto \sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}k_3^{\frac{n+1+2h_3-2}{2}}\exp\left(-\frac{1}{2}k_3(B^2e_3^2+1)\right)$$

Let $\widetilde{V_4}^{-1} = (B^2e_3^2 + 1)$. Then, using (A.1.15) and (A.1.4), we have the distribution of $k_4|y_3, y_2, y_1$ as follows:

$$\pi(k_4|y_3, y_2, y_1) \propto \int k_4^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_4\Big) {}_0F_1\Big(\frac{n}{2}; \frac{1}{4}\rho^2 k_3 k_4\Big) \exp\Big(-\frac{1}{2}\rho^2 k_3\Big) \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}} \times \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} k_3^{\frac{n+1+2h_3-2}{2}} \exp\Big(-\frac{1}{2}k_3\widetilde{V_4}^{-1}\Big) dk_3 \tag{A.1.17}$$

Taking this integral with respect to k_3 we get:

$$\pi(k_4|y_3, y_2, y_1) \propto k_4^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_4\Big) \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} {}_1F_1\Big(\frac{n+1+2h_3}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_4(\widetilde{V_4}^{-1} + \rho^2)^{-1}\Big) \Gamma\Big(\frac{n+1+2h_3}{2}\Big)(2S_4)^{\frac{n+1+2h_3}{2}} \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}$$

where $S_4 = (\widetilde{V_4}^{-1} + \rho^2)^{-1} = (B^2e_3^2 + 1 + \rho^2)^{-1}$. Let c_4 be the normalising constant, that is:

$$c_4 = \int k_4^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_4\Big) \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} {}_1F_1\Big(\frac{n+1+2h_3}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_4(\widetilde{V_4}^{-1} + \rho^2)^{-1}\Big) \Gamma\Big(\frac{n+1+2h_3}{2}\Big)(2S_4)^{\frac{n+1+2h_3}{2}} \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}} dk_4$$

Thus we get:

$$c_4 = \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} {}_2F_1\Big(\frac{n+1+2h_3}{2}, \frac{n}{2}; \frac{n}{2}; \rho^2 S_4\Big)\Gamma\Big(\frac{n+1+2h_3}{2}\Big)(2S_4)^{\frac{n+1+2h_3}{2}}$$

Using (A.1.10) and (A.1.11), this simplifies to:

$$c_4 = \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3}(1 - \rho^2 S_4)^{-\frac{n+1+2h_3}{2}}\Gamma\Big(\frac{n+1+2h_3}{2}\Big)(2S_4)^{\frac{n+1+2h_3}{2}}$$

Thus,

$$\pi(k_4|y_3,y_2,y_1)=\frac{1}{c_4}{k_4}^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_4\Big)\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}{}_1F_1\Big(\frac{n+1+2h_3}{2};\frac{n}{2};\frac{1}{2}\rho^2k_4({\widetilde{V}_4}^{-1}+\rho^2)^{-1}\Big)$$
$$\Gamma\Big(\frac{n+1+2h_3}{2}\Big)(2S_4)^{\frac{n+1+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}$$

Therefore the likelihood for $t=4$ is as follows:

$$L(y_4|y_3,y_2,y_1)=\int\pi(y_4|k_4,y_3,y_2,y_1)\pi(k_4|y_3,y_2,y_1)dk_4$$

Thus we have:

$$L(y_4|y_3,y_2,y_1)=\int(2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_4}{k_4}^{\frac{n+1-2}{2}}\exp\Big(-\frac{1}{2}k_4(B^2e_4^2+1)\Big)\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\Gamma\Big(\frac{n+1+2h_3}{2}\Big)$$
$${}_1F_1\Big(\frac{n+1+2h_3}{2};\frac{n}{2};\frac{1}{2}k_4\rho^2S_4\Big)(2S_4)^{\frac{n+1+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_4$$

This is similar to $t=3$ therefore we have:

$$L(y_4|y_3,y_2,y_1)=(2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_4}\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\frac{\Gamma\Big(\frac{n+1+2h_3}{2}\Big)}{(B^2e_4^2+1)^{\frac{n+1}{2}}}(2S_4)^{\frac{n+1+2h_3}{2}}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+1}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_4$$

and the likelihood for any t is:

$$L(y_t|y_{1:t-1})=(2\pi)^{-\frac{1}{2}}\sqrt{B^2}\frac{1}{c_t}\sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}\frac{\Gamma\Big(\frac{n+1+2h_{t-1}}{2}\Big)}{(B^2e_t^2+1)^{\frac{n+1}{2}}}(2S_t)^{\frac{n+1+2h_{t-1}}{2}}\frac{2^{\frac{n+1}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+1}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_t$$

where for $t\geq 4$:

$$\delta_t = \left((1-\rho^2 S_t)^{-1} S_t \rho^2 (\widetilde{V}_{t-1}^{-1} + \rho^2)^{-1}\right)$$

$$Z_t = (B^2 e_t^2 + 1)^{-1} S_t \rho^2$$

$$\hat{C}_t = {}_2F_1\left(\frac{n+1+2h_{t-1}}{2}, \frac{n+1}{2}; \frac{n}{2}; Z_t\right)$$

$$\widetilde{V}_t^{-1} = 1 + B^2 e_{t-1}^2$$

$$S_t = (B^2 e_{t-1}^2 + 1 + \rho^2)^{-1} = (\widetilde{V}_t^{-1} + \rho^2)^{-1}$$

$$c_t = \sum_{h_{t-1}=0}^{\infty} \widetilde{C}_{t-1,h_{t-1}} (1-\rho^2 S_t)^{-\frac{n+1+2h_{t-1}}{2}} \Gamma\left(\frac{n+1+2h_{t-1}}{2}\right) (2S_t)^{\frac{n+1+2h_{t-1}}{2}}$$

$$\widetilde{C}_{t-1,h_{t-1}} =$$

$$\sum_{h_{t-2}=0}^{\infty} \widetilde{C}_{t-2,h_{t-2}} \Gamma\left(\frac{n+1+2h_{t-2}}{2}\right) \frac{[(n+1)/2 + h_{t-2}]_{h_{t-1}}}{[n/2]_{h_{t-1}}} \left(\frac{1}{2}\rho^2 S_{t-1}\right)^{h_{t-1}} \frac{(2S_{t-1})^{\frac{n+1+2h_{t-2}}{2}}}{h_{t-1}!}$$

□

A.1.3 Proof of Proposition 3.3.2

Proof. Combining the prior density for k_1 in (A.1.1) with the transition equation in (A.1.4) and the likelihood, we get:

$$\begin{aligned} \pi(k_1|k_{2:T}, y_{1:T}) &\propto |k_1|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_2^{-1}k_1\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_1 k_2\right) \\ &= |k_1|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_2^{-1}k_1\right) \sum_{h=0}^{\infty} \left(C_{1,h}|k_1|^h\right) \end{aligned} \tag{A.1.18}$$

with $C_{1,h} = \frac{1}{h!}\frac{1}{[n/2]_h}\left(\frac{1}{4}\rho^2 k_2\right)^h$.

The integral of (A.1.18) with respect to k_1 is proportional to:

$${}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_2 S_2\right)$$

and therefore:

$$\begin{aligned} &\pi(k_2|k_{3:T}, y_{1:T}) \\ &\propto |k_2|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_3^{-1}k_2\right) {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_2 S_2\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_3 k_2\right) \end{aligned} \tag{A.1.19}$$

where we have used that $S_3^{-1} = 1 + B^2 e_2^2 + \rho^2$. Combining the series we get that:

$$
\begin{aligned}
&{}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_2 S_2\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_3 k_2\right) = \\
&\left(\sum_{h_1=0}^{\infty} \frac{[(n+1)/2]_{h_1}}{[n/2]_{h_1}} \frac{(\frac{1}{2}\rho^2 S_2)^{h_1} k_2^{h_1}}{h_1!}\right)\left(\sum_{h_2=0}^{\infty} \frac{1}{h_2!} \frac{1}{[n/2]_{h_2}} \left(\frac{1}{4}\rho^2 k_3\right)^{h_2} k_2^{h_2}\right)
\end{aligned}
\tag{A.1.20}
$$

By making the change of variables $h = h_1 + h_2$ we get that (A.1.20) can be written as:

$$
\sum_{h=0}^{\infty} \sum_{h_2=0}^{h} \left(\left(\frac{[(n+1)/2]_{h-h_2}}{[n/2]_{h-h_2}} \frac{(\frac{1}{2}\rho^2 S_2)^{h-h_2}}{(h-h_2)!}\right) \frac{1}{h_2!} \frac{1}{[n/2]_{h_2}} \left(\frac{1}{4}\rho^2\right)^{h_2} k_3^{h_2}\right) k_2^h = \sum_{h=0}^{\infty} C_{2,h} k_2^h \tag{A.1.21}
$$

where:

$$
C_{2,h} = \sum_{h_2=0}^{h} \widetilde{C}_{2,h-h_2} \frac{1}{h_2!} \frac{1}{[n/2]_{h_2}} \left(\frac{1}{4}\rho^2\right)^{h_2} k_3^{h_2}
$$

and $\widetilde{C}_{2,h-h_2}$ has been defined in proposition 3.1 as:

$$
\widetilde{C}_{2,h-h_2} = \frac{[(n+1)/2]_{h-h_2}}{[n/2]_{h-h_2}} \frac{(\frac{1}{2}\rho^2 S_2)^{h-h_2}}{(h-h_2)!}
$$

Using (A.1.21) we obtain that:

$$
\pi(k_2 | k_{3:T}, y_{1:T}) \propto |k_2|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2} S_3^{-1} k_2\right) \sum_{h=0}^{\infty} \left(C_{2,h} k_2^h\right) \tag{A.1.22}
$$

as we wanted to prove.

The integral of (A.1.22) with respect to k_2 is proportional to:

$$
\sum_{h=0}^{\infty} \left(C_{2,h} \frac{\Gamma\left(\frac{n+1+2h}{2}\right)}{(S_3^{-1}/2)^{\frac{n+1+2h}{2}}}\right) = \sum_{h=0}^{\infty} \left(\sum_{h_2=0}^{h} \widetilde{C}_{2,h-h_2} \frac{1}{h_2!} \frac{1}{[n/2]_{h_2}} \left(\frac{1}{4}\rho^2\right)^{h_2} k_3^{h_2}\right) \frac{\Gamma\left(\frac{n+1+2h}{2}\right)}{(S_3^{-1}/2)^{\frac{n+1+2h}{2}}} \tag{A.1.23}
$$

Making the change of variables $h_1 = h - h_2$, equation (A.1.23) can be written as:

$$
\sum_{h_1=0}^{\infty} \sum_{h_2=0}^{\infty} \left(\widetilde{C}_{2,h_1} \frac{1}{h_2!} \frac{1}{[n/2]_{h_2}} \left(\frac{1}{4}\rho^2\right)^{h_2} k_3^{h_2}\right) \frac{\Gamma\left(\frac{n+1}{2} + h_1 + h_2\right)}{(S_3^{-1}/2)^{\frac{n+1}{2} + h_1 + h_2}} \tag{A.1.24}
$$

Note that $\Gamma\big(\frac{n+1}{2}+h_1+h_2\big)=\Gamma\big(\frac{n+1+2h_1}{2}\big)\big[\frac{n+1+2h_1}{2}\big]_{h_2}$. Then (A.1.24) can be written as:

$$\sum_{h_2=0}^{\infty}\sum_{h_1=0}^{\infty}\widetilde{C}_{2,h_1}\Gamma\bigg(\frac{n+1+2h_1}{2}\bigg)\frac{\big[(n+1)/2+h_1\big]_{h_2}}{[n/2]_{h_2}}\bigg(\frac{1}{2}\rho^2S_3\bigg)^{h_2}\frac{1}{h_2!}(2S_3)^{\frac{n+1+2h_1}{2}}k_3^{h_2} \tag{A.1.25}$$

Using the definition of $\widetilde{C}_{3,h_2}$ in proposition 3.1, we can write (A.1.25) as:

$$\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}$$

Recall that the transition density is in (A.1.4). Therefore, we have:

$$\pi(k_3|k_{4:T},y_{1:T})\propto\bigg(\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}\bigg){}_0F_1\bigg(\frac{n}{2};\frac{1}{4}\rho^2k_3k_4\bigg)|k_3|^{\frac{n+1-2}{2}}\exp\bigg(-\frac{1}{2}S_4^{-1}k_3\bigg)$$

with $S_4^{-1}=1+B^2e_3^2+\rho^2$. As before, we can multiply the two series as follows:

$$\bigg(\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}\bigg){}_0F_1\bigg(\frac{n}{2};\frac{1}{4}\rho^2k_3k_4\bigg)=\bigg(\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}\bigg)\bigg(\sum_{h_3=0}^{\infty}\frac{1}{[n/2]_{h_3}}\bigg(\frac{1}{4}\rho^2k_3\bigg)^{h_3}k_4^{h_3}\frac{1}{h_3!}\bigg)$$
$$=\sum_{h=0}^{\infty}\sum_{h_3=0}^{h}|k_3|^h\widetilde{C}_{3,h-h_3}\frac{1}{[n/2]_{h_3}}\bigg(\frac{1}{4}\rho^2\bigg)^{h_3}k_4^{h_3}\frac{1}{h_3!}=\sum_{h=0}^{\infty}|k_3|^hC_{3,h}$$

where

$$C_{3,h}=\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h-h_3}\frac{1}{[n/2]_{h_3}}\bigg(\frac{1}{4}\rho^2\bigg)^{h_3}\frac{k_4^{h_3}}{h_3!}$$

and therefore, $\pi(k_3|k_{4:T},y_{1:T})$ can be written as:

$$\pi(k_3|k_{4:T},y_{1:T})\propto|k_3|^{\frac{n+1-2}{2}}\exp\bigg(-\frac{1}{2}S_4^{-1}k_3\bigg)\sum_{h=0}^{\infty}|k_3|^hC_{3,h} \tag{A.1.26}$$

as we wanted to prove.

Since $\pi(k_3|k_{4:T},y_{1:T})$ in (A.1.26) and $\pi(k_2|k_{3:T},y_{1:T})$ in (A.1.22) have the same structure, and, since the transition density of k_t is always the same, we get analogous results for any $t<T$, as we wanted to prove. For $t=T$ the only difference is that there is no

transition density from k_T to k_{T+1}. For this reason we do not need to multiply two series, and hence $C_{T,h} = \widetilde{C}_{T,h}$ and $S_{T+1} = (1 + B^2 e_T^2)^{-1}$

□

A.1.4 Proof of Proposition 3.3.3

Proof. We need to integrate $\pi(k_{1:T})\pi(y_{1:T}|k_{1:T})$ with respect to k_T first. The terms that depend on k_T are the following:

$$\begin{aligned}\exp\Big(-\frac{1}{2}e_T^2 B^2 k_T\Big)|k_T|^{\frac{1}{2}}|k_T|^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_T\Big){}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_T k_{T-1}\Big) = \\ \exp\Big(-\frac{1}{2}S_{T+1}^{-1}k_T\Big)|k_T|^{\frac{n+1-2}{2}}\sum_{h=0}^{\infty}a_{T,h}|k_T|^h\end{aligned} \tag{A.1.27}$$

with $a_{T,h} = \frac{1}{h!}\frac{1}{[n/2]_h}\big(\frac{1}{4}\rho^2 k_{T-1}\big)^h$. This proves the result for $s = 0$. The integral of (A.1.27) with respect to k_T is proportional to:

$${}_1F_1\Big(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}\rho^2 k_{T-1}S_{T+1}\Big)$$

Therefore, the terms that depend on k_{T-1} in $\pi(k_{1:T})\pi(y_{1:T}|k_{1:T})$ after integrating out k_T are the following:

$$|k_{T-1}|^{\frac{n+1-2}{2}}\exp\Big(-\frac{1}{2}S_T^{-1}k_{T-1}\Big){}_1F_1\Big(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}\rho^2 k_{T-1}S_{T+1}\Big){}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_{T-1}k_{T-2}\Big) \tag{A.1.28}$$

Equation (A.1.28) has the product of two series, that can be written as:

$$\begin{aligned}&{}_1F_1\Big(\frac{n+1}{2};\frac{n}{2};\frac{1}{2}\rho^2 k_{T-1}S_{T+1}\Big){}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_{T-1}k_{T-2}\Big) = \\ &= \Big(\sum_{h_T=0}^{\infty}\frac{[(n+1)/2]_{h_T}}{[n/2]_{h_T}}\frac{(\frac{1}{2}\rho^2 S_{T+1})^{h_T}k_{T-1}^{h_T}}{h_T!}\Big)\Big(\sum_{h_{T-1}=0}^{\infty}\frac{1}{(h_{T-1})!}\frac{1}{[n/2]_{h_{T-1}}}\Big(\frac{1}{4}\rho^2 k_{T-2}\Big)^{h_{T-1}}k_{T-1}^{h_{T-1}}\Big)\end{aligned} \tag{A.1.29}$$

Making the change of variables $h = h_T + h_{T-1}$ we get that (A.1.29) is equal to:

$$\sum_{h=0}^{\infty}\Bigg(\sum_{h_{T-1}=0}^{h}\frac{[(n+1)/2]_{h-h_{T-1}}}{[n/2]_{h-h_{T-1}}}\frac{(\frac{1}{2}\rho^2 S_{T+1})^{h-h_{T-1}}}{(h-h_{T-1})!}\frac{1}{(h_{T-1})!}\frac{1}{[n/2]_{h_{T-1}}}\left(\frac{1}{4}\rho^2\right)^{h_{T-1}}k_{T-2}^{h_{T-1}}\Bigg)k_{T-1}^{h}=$$
$$\sum_{h=0}^{\infty}a_{T-1,h}k_{T-1}^{h}$$

where:

$$a_{T-1,h}=\sum_{h_{T-1}=0}^{h}\left(\frac{[(n+1)/2]_{h-h_{T-1}}}{[n/2]_{h-h_{T-1}}}\frac{(\frac{1}{2}\rho^2 S_{T+1})^{h-h_{T-1}}}{(h-h_{T-1})!}\frac{1}{(h_{T-1})!}\frac{1}{[n/2]_{h_{T-1}}}\left(\frac{1}{4}\rho^2\right)^{h_{T-1}}k_{T-2}^{h_{T-1}}\right)$$

which can be written as:

$$a_{T-1,h}=\sum_{h_{T-1}=0}^{h}\widetilde{a}_{T-1,h-h_{T-1}}\frac{1}{(h_{T-1})!}\frac{1}{[n/2]_{h_{T-1}}}\left(\frac{1}{4}\rho^2\right)^{h_{T-1}}k_{T-2}^{h_{T-1}}$$

and:

$$\widetilde{a}_{T-1,h}=\frac{[(n+1)/2]_h}{[n/2]_h}\frac{(\frac{1}{2}\rho^2 S_{T+1})^h}{h!}$$

Therefore, $\pi(k_{T-1}|k_{1:T-2},y_{1:T})$ which is given by (A.1.28), can be written as:

$$\pi(k_{T-1}|k_{1:T-2},y_{1:T})\propto|k_{T-1}|^{\frac{n+1-2}{2}}\exp\left(-\frac{1}{2}S_T^{-1}k_{T-1}\right)\sum_{h=0}^{\infty}\left(a_{T-1,h}k_{T-1}^{h}\right)\qquad(A.1.30)$$

which proves the result for $s=1$.

The integral of (A.1.30) with respect to k_{T-1} gives:

$$\begin{aligned}&\sum_{h=0}^{\infty}\left(a_{T-1,h}\frac{\Gamma\left(\frac{n+1+2h}{2}\right)}{(S_T^{-1}/2)^{\frac{n+1+2h}{2}}}\right)=\\&=\sum_{h=0}^{\infty}\sum_{h_{T-1}=0}^{h}\widetilde{a}_{T-1,h-h_{T-1}}\frac{1}{(h_{T-1})!}\frac{1}{[n/2]_{h_{T-1}}}\left(\frac{1}{4}\rho^2\right)^{h_{T-1}}k_{T-2}^{h_{T-1}}\frac{\Gamma\left(\frac{n+1+2h}{2}\right)}{(S_T^{-1}/2)^{\frac{n+1+2h}{2}}}\end{aligned}\qquad(A.1.31)$$

Making a change of variables $h = h_T + h_{T-1}$, equation (A.1.31) can be written as:

$$\sum_{h_T=0}^{\infty}\sum_{h_{T-1}=0}^{\infty}\left(\widetilde{a}_{T-1,h_T}\frac{1}{(h_{T-1})!}\frac{1}{[n/2]_{h_{T-1}}}\left(\frac{1}{4}\rho^2\right)^{h_{T-1}}k_{T-2}^{h_{T-1}}\right)\frac{\Gamma\left(\frac{n+1}{2}+h_T+h_{T-1}\right)}{(S_T^{-1}/2)^{\frac{n+1}{2}+h_T+h_{T-1}}} \quad \text{(A.1.32)}$$

Noting that $\Gamma\left(\frac{n+1}{2}+h_T+h_{T-1}\right) = \Gamma\left(\frac{n+1}{2}+h_T\right)\left[\frac{n+1}{2}+h_T\right]_{h_{T-1}}$, (A.1.32) can be written as:

$$\sum_{h_{T-1}=0}^{\infty}\left(\sum_{h_T=0}^{\infty}\widetilde{a}_{T-1,h_T}\Gamma\left(\frac{n+1}{2}+h_T\right)\frac{\left[(n+1)/2+h_T\right]_{h_{T-1}}}{[n/2]_{h_{T-1}}}\frac{\left(\frac{1}{2}\rho^2 S_T\right)^{h_{T-1}}}{(h_{T-1)}!}(2S_T)^{\frac{n+1+2h_T}{2}}\right)k_{T-2}^{h_{T-1}}$$
$$= \sum_{h_{T-1}=0}^{\infty}\widetilde{a}_{T-2,h_{T-1}}k_{T-2}^{h_{T-1}} \quad \text{(A.1.33)}$$

where:

$$\widetilde{a}_{T-2,h_{T-1}} = \sum_{h_T=0}^{\infty}\widetilde{a}_{T-1,h_T}\Gamma\left(\frac{n+1}{2}+h_T\right)\frac{\left[(n+1)/2+h_T\right]_{h_{T-1}}}{[n/2]_{h_{T-1}}}\frac{\left(\frac{1}{2}\rho^2 S_T\right)^{h_{T-1}}}{(h_{T-1})!}(2S_T)^{\frac{n+1+2h_T}{2}}$$

Therefore, we have that the integral of (A.1.30) with respect to k_{T-1} gives (A.1.33).Therefore, collecting the terms that depend on k_{T-2} we have that:

$$\pi(k_{T-2}|k_{1:(T-3)}, y_{1:T}) \propto$$
$$|k_{T-2}|^{\frac{n+1-2}{2}}\exp\left(-\frac{1}{2}S_{T-1}^{-1}k_{T-2}\right)\left(\sum_{h_{T-1}=0}^{\infty}\widetilde{a}_{T-2,h_{T-1}}k_{T-2}^{h_{T-1}}\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2 k_{T-2}k_{T-3}\right) \quad \text{(A.1.34)}$$

Equation (A.1.34) depends on the product of two series, which can be written as follows:

$$\Bigg(\sum_{h_{T-1}=0}^{\infty}\widetilde{a}_{T-2,h_{T-1}}k_{T-2}^{h_{T-1}}\Bigg){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2k_{T-2}k_{T-3}\right)=$$

$$\Bigg(\sum_{h_{T-1}=0}^{\infty}\widetilde{a}_{T-2,h_{T-1}}k_{T-2}^{h_{T-1}}\Bigg)\Bigg(\sum_{h_{T-2}=0}^{\infty}\frac{\left(\frac{1}{4}\rho^2k_{T-2}k_{T-3}\right)^{h_{T-2}}}{(h_{T-2})!}\frac{1}{[n/2]_{h_{T-2}}}\Bigg)=$$

$$\sum_{h=0}^{\infty}\Bigg(\sum_{h_{T-2}=0}^{h}\widetilde{a}_{T-2,h-h_{T-2}}\frac{1}{(h_{T-2})!}\frac{1}{[n/2]_{h_{T-2}}}\left(\frac{1}{4}\rho^2k_{T-3}\right)^{h_{T-2}}\Bigg)k_{T-2}^{h}=\sum_{h=0}^{\infty}a_{T-2,h}k_{T-2}^{h}$$

where:

$$a_{T-2,h}=\sum_{h_{T-2}=0}^{h}\widetilde{a}_{T-2,h-h_{T-2}}\frac{1}{(h_{T-2})!}\frac{1}{[n/2]_{h_{T-2}}}\left(\frac{1}{4}\rho^2k_{T-3}\right)^{h_{T-2}}$$

Therefore, we can write (A.1.34) as:

$$\pi(k_{T-2}|k_{1:(T-3)},y_{1:T})\propto|k_{T-2}|^{\frac{n+1-2}{2}}\exp\left(-\frac{1}{2}S_{T-1}^{-1}k_{T-2}\right)\sum_{h=0}^{\infty}a_{T-2,h}k_{T-2}^{h}\qquad\text{(A.1.35)}$$

which proves the result for $s=2$.

Because $\pi(k_{T-2}|k_{1:(T-3)},y_{1:T})$ in (A.1.35) and $\pi(k_{T-1}|k_{1:T-2},y_{1:T})$ in (A.1.30) have the same structure, and because the transition density is always the same, we can conclude the result is proven for any $s=0,\ldots,T-2$. For $s=T-1$ there is no transition density from k_0 to k_1, therefore there is no need to multiply two series, so we get $a_{1,h}=\widetilde{a}_{1,h}$ and $S_2=(1+B^2e_1^2)^{-1}$.

□

A.1.5 Proof of Proposition 3.3.4

Proof. To find $\pi(k_t|y_{1:T})$ we need to integrate $\pi(k_{1:T})\pi(y_{1:T}|k_{1:T})$ with respect to $k_{1:(t-1)}$ and with respect to $k_{(t+1):T}$. From the proofs of propositions 3.2 and 3.3, we have that when $2\leq t<(T-1)$:

$$\int\int \pi(k_{1:T})\pi(y_{1:T}|k_{1:T})dk_{1:(t-1)}dk_{(t+2):T} \propto |k_t|^{\frac{n+1-2}{2}} \exp\Big(-\frac{1}{2}S_{t+1}^{-1}k_t\Big)\Big(\sum_{h=0}^{\infty} C_{t,h}|k_t|^h\Big)$$
$$\times |k_{t+1}|^{\frac{n+1-2}{2}} \exp\Big(-\frac{1}{2}S_{t+2}^{-1}k_{t+1}\Big)\Big(\sum_{h=0}^{\infty} a_{t+2,h}\frac{\Gamma\big(\frac{n+1}{2}+h\big)}{(S_{t+3}^{-1}/2)^{\frac{n+1+2h}{2}}}\Big) \tag{A.1.36}$$

In the proof of proposition 3.3, it is shown that:

$$\sum_{h=0}^{\infty} a_{t+2,h}\frac{\Gamma\big(\frac{n+1}{2}+h\big)}{(S_{t+3}^{-1}/2)^{\frac{n+1+2h}{2}}} = \sum_{h=0}^{\infty}\widetilde{a}_{t+1,h}k_{t+1}^h$$

Therefore (A.1.36) can be written as:

$$\pi(k_t,k_{t+1}|y_{1:T}) \propto |k_t|^{\frac{n+1-2}{2}} \exp\Big(-\frac{1}{2}S_{t+1}^{-1}k_t\Big)\Big(\sum_{h=0}^{\infty} C_{t,h}|k_t|^h\Big)$$
$$\times |k_{t+1}|^{\frac{n+1-2}{2}} \exp\Big(-\frac{1}{2}S_{t+2}^{-1}k_{t+1}\Big)\Big(\sum_{h=0}^{\infty}\widetilde{a}_{t+1,h}k_{t+1}^h\Big) \tag{A.1.37}$$

The product of the two series can be written as:

$$\Big(\sum_{h=0}^{\infty} C_{t,h}|k_t|^h\Big)\Big(\sum_{h=0}^{\infty}\widetilde{a}_{t+1,h}k_{t+1}^h\Big) =$$
$$\sum_{h_{t+1}=0}^{\infty}\sum_{h=0}^{\infty}\sum_{h_t=0}^{h}\widetilde{C}_{t,h-h_t}\frac{1}{[n/2]_{h_t}}\Big(\frac{1}{4}\rho^2\Big)^{h_t}\frac{(k_{t+1})^{h_t+h_{t+1}}}{h_t!}|k_t|^h\widetilde{a}_{t+1,h_{t+1}} \tag{A.1.38}$$

where neither $\widetilde{a}_{t+1,h_{t+1}}$ nor $\widetilde{C}_{t,h-h_t}$ depend on k_{t+1}. Therefore, we can integrate out k_{t+1} from (A.1.37) using (A.1.38) to obtain:

$$\pi(k_t|y_{1:T}) \propto |k_t|^{\frac{n+1-2}{2}} \exp\Big(-\frac{1}{2}S_{t+1}^{-1}k_t\Big)\Big(\sum_{h=0}^{\infty}\widetilde{D}_{t,h}|k_t|^h\Big)$$

where

$$\widetilde{D}_{t,h} = \sum_{h_t=0}^{h}\sum_{h_{t+1}=0}^{\infty}\widetilde{C}_{t,h-h_t}\frac{1}{[n/2]_{h_t}}\Big(\frac{1}{4}\rho^2\Big)^{h_t}\frac{1}{h_t!}\frac{\Gamma\big(\frac{n+1}{2}+h_t+h_{t+1}\big)}{(S_{t+2}^{-1}/2)^{\frac{n+1}{2}+h_t+h_{t+1}}}\widetilde{a}_{t+1,h_{t+1}}$$

as we wanted to prove.

In the case $t = T-1$, expression (A.1.37) becomes:

$$\pi(k_{T-1}, k_T|y_{1:T}) \propto |k_{T-1}|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_T^{-1}k_{T-1}\right)\left(\sum_{h=0}^{\infty} C_{T-1,h}|k_{T-1}|^h\right)|k_T|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_{T+1}^{-1}k_T\right) \tag{A.1.39}$$

Thus, in this case we only have one series, not the product of two. Integrating with respect to k_T we get:

$$\pi(k_{T-1}|y_{1:T}) \propto |k_{T-1}|^{\frac{n+1-2}{2}} \exp\left(-\frac{1}{2}S_T^{-1}k_{T-1}\right)\sum_{h=0}^{\infty} \widetilde{D}_{T-1,h}|k_{T-1}|^h$$

with

$$\widetilde{D}_{T-1,h} = \sum_{h_{T-1}=0}^{h} \widetilde{C}_{T-1,h-h_{T-1}} \frac{1}{[n/2]_{h_{T-1}}}\left(\frac{1}{4}\rho^2\right)^{h_{T-1}} \frac{1}{(h_{T-1})!}\frac{\Gamma\left(\frac{n+1}{2}+h_{T-1}\right)}{(S_{T+1}^{-1}/2)^{\frac{n+1}{2}+h_{T-1}}}$$

as we wanted to prove.

□

A.1.6 Proof of Local Scale Model Likelihood

To facilitate the reading we do not explicitly write x_t as a conditioning argument. Given that we have a gamma distribution for the initial condition (3.2.2) and a Gaussian error term, we have that the joint density (y_1, h_1, ν_1) is :

$$\pi(y_1, h_1, \nu_1) = \frac{1}{\sqrt{2\pi}}(h_1)^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y_1 - x_1\beta)^2 h_1\right) f(h_1|S_1)\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\nu_1^{\alpha_1-1}(1-\nu_1)^{\alpha_2-1}$$

where $f(h_1|S_1)$ is the density of the initial condition given as:

$$f(h_1|S_1) = h_1^{\frac{\nu}{2}-1} \exp\left(-\frac{h_1}{2S_1}\right)\frac{1}{\Gamma(\nu/2)(2S_1)^{\frac{\nu}{2}}} \tag{A.1.40}$$

The volatility process is represented by a non stationary process as in (3.2.1). We make a change of variables from (y_1, h_1, ν_1) to (y_1, Z, h_2) where $Z = h_1 - \lambda h_2$, and $v_1 = \frac{h_2\lambda}{h_1}$. The

Jacobian of this transformation is $\lambda/(Z+\lambda h_2)$. Therefore $\pi(y_1, Z, h_2)$ can be written as:

$$\pi(y_1, Z, h_2) = \frac{1}{\sqrt{2\pi}}(Z+\lambda h_2)^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y_1 - x_1\beta)^2(Z+\lambda h_2)\right)(Z+\lambda h_2)^{\frac{\nu}{2}-1} \exp\left(-\frac{1}{2S_1}(Z+\lambda h_2)\right)$$
$$\times \left(\frac{(Z+\lambda h_2)}{\lambda}\right)^{-1} \frac{1}{\Gamma(\nu/2)(2S_1)^{\frac{\nu}{2}}} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \left(\frac{h_2\lambda}{Z+\lambda h_2}\right)^{\alpha_1-1} \left(\frac{Z}{Z+\lambda h_2}\right)^{\alpha_2-1}$$

which simplifies to:

$$\pi(y_1, Z, h_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left((y_1 - x_1\beta)^2 + \frac{1}{S_1}\right)(Z+\lambda h_2)\right) \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{(Z+\lambda h_2)^{\frac{\nu}{2}+\frac{1}{2}-(\alpha_1+\alpha_2)}}{\Gamma(\nu/2)(2S_1)^{\frac{\nu}{2}}} \lambda_1^{\alpha} Z^{\alpha_2-1} h_2^{\alpha_1}$$

Note that for mathematical convenience, α_1 is restricted as $\alpha_1 = \frac{\nu}{2}$ and $\alpha_2 = \frac{1}{2}$. Therefore, $\frac{\nu}{2} + \frac{1}{2} - (\alpha_1 + \alpha_2) = 0$, and $\pi(Z|y_1, h_2)$ is a gamma distribution. Using the properties of the gamma distribution, we can integrate over the state variable Z:

$$\begin{aligned}\pi(y_1, h_2) &= \int \pi(y_1, Z, h_2) dZ \\ &= \frac{\Gamma(\alpha_2)}{\sqrt{2\pi}} \frac{\exp\left(-\frac{\lambda h_2}{2}\left((y_1 - x_1\beta)^2 + \frac{1}{S_1}\right)\right)}{\left(2\left((y_1 - x_1\beta)^2 + \frac{1}{S_1}\right)^{-1}\right)^{-\alpha_2}} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\lambda_1^{\alpha} h_2^{\nu/2-1}}{\Gamma(\nu/2)(2S_1)^{\frac{\nu}{2}}} \qquad \text{(A.1.41)}\end{aligned}$$

From equation (A.1.41) we can see that $h_2|y_1$ is a gamma distribution with parameters $\left(\frac{\nu}{2}, 2S_2\right)$, where $S_2 = \left((y_1 - x_1\beta)^2 + \frac{1}{S_1}\right)^{-1}\frac{1}{\lambda}$. Let $f(h_2|S_2)$ be defined as in (A.1.40), that is, the density of a gamma distribution:

$$f(h_2|S_2) = h_2^{\frac{\nu}{2}-1} \exp\left(-\frac{h_2}{2S_2}\right) \frac{1}{\Gamma(\nu/2)(2S_2)^{\frac{\nu}{2}}}. \qquad \text{(A.1.42)}$$

Then equation (A.1.41) can be written as follows:

$$\pi(y_1, h_2) = \frac{\Gamma(\alpha_2)}{\Gamma(\alpha_2)} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)} \frac{\lambda^{\alpha_1}}{\Gamma(\nu/2)(2S_1)^{\frac{\nu}{2}}} \frac{1}{\sqrt{2\pi}} \left(2\left((y_1 - x_1\beta)^2 + \frac{1}{S_1}\right)^{-1}\right)^{\alpha_2} f(h_2|S_2)\Gamma(\nu/2)(2S_2)^{\frac{\nu}{2}}$$

Therefore, $h_2|y_1$ is a gamma distribution, such that $\pi(h_2|y_1) = f(h_2|S_2)$, which is defined in (A.1.42).

From these derivations we can get the likelihood as follows. First, for $t = 1$, we have

that

$$\pi(y_1|h_1) = \frac{1}{(\sqrt{2\pi})} h_1^{\frac{1}{2}} \exp\left(-\frac{1}{2} h_1 (y_1 - x_1\beta)^2 \right)$$

and the initial condition for h_1 is a gamma distribution given in (A.1.40). Therefore, $\pi(y_1)$ is a student-t and we have:

$$\pi(y_1) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)} \lambda^{\alpha_1} \left(\frac{S_2}{S_1}\right)^{\alpha_1} \frac{1}{\sqrt{2\pi}} \left(2\left((y_1 - x_1\beta)^2 + \frac{1}{S_1}\right)^{-1}\right)^{\alpha_2} \tag{A.1.43}$$

For $t = 2$, $\pi(y_2|h_2)$ is also a normal. Thus the conditional distribution for the second observation given h_2 is as follows:

$$\pi(y_2|h_2) = \frac{1}{(\sqrt{2\pi})} h_2^{\frac{1}{2}} \exp\left(-\frac{1}{2} h_2 (y_2 - x_2\beta)^2 \right)$$

and $\pi(h_2|y_1)$ is the gamma distribution defined in (A.1.42). Therefore, we have the same structure as in $t = 1$, and using the properties of the gamma distribution , we get that the likelihood $\pi(y_2|y_1)$ is a student-t as follows:

$$\pi(y_2|y_1) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)} \lambda^{\alpha_1} \left(\frac{S_3}{S_2}\right)^{\alpha_1} \frac{1}{\sqrt{2\pi}} \left(2\left((y_2 - x_2\beta)^2 + \frac{1}{S_2}\right)^{-1}\right)^{\alpha_2} \tag{A.1.44}$$

where $S_3 = \left((y_2 - x_2\beta)^2 + \frac{1}{S_2}\right)^{-1} \frac{1}{\lambda}$.

Because the kernels are the same for $t = 1$ and for $t = 2$, then we have proved it for every t.

A.2 Appendices to Chapter 4

A.2.1 Proof of Lemma 4.1

To derive the likelihood we will make use of the following lemma.

Lemma A.2.1.

$$\int |K|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2} AK \right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4} BK\right) dK = \Gamma\left(\frac{n+r}{2}\right) \left|\frac{1}{2}A\right|^{-\frac{n+r}{2}} {}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2} BA^{-1}\right)$$

where ${}_0F_1(.)$ *and* ${}_1F_1(.)$ *are hypergeometric series.*

whose proof is as below:

Proof. The integral is a gamma multiplied by a hypergeometric function. Therefore, the integral is very standard so we can use the properties of hypergeometric functions. We apply Theorem 7.3.4 in Muirhead (2005) to get the result, thus, we transform the functions by applying a change of variables. Let $X = \frac{1}{4}BK$ such that $K = 4XB^{-1}$ thus we have:

$$ {}_0F_1\left(\frac{n}{2}; \frac{1}{4}BK\right) = {}_0F_1\left(\frac{n}{2}; X\right) $$

Then;

$$ \int |X|^{\frac{n+r-2}{2}} |4B^{-1}|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2}4AXB^{-1}\right) {}_0F_1\left(\frac{n}{2}; X\right) dK $$

We use the Jacobian $dK = |4B^{-1}|dX$ to integrate with respect to X;

$$ \int |X|^{\frac{n+r-2}{2}} \exp(-2XB^{-1}A) {}_0F_1\left(\frac{n}{2}; X\right) dX |4B^{-1}|^{\frac{n+r}{2}} $$

This integral is the same as in the theorem, therefore, when we integrate out K we get the following:

$$ \int |X|^{\frac{n+r-2}{2}} \exp(-X2B^{-1}A) {}_0F_1\left(\frac{n}{2}; X\right) dX |4B^{-1}|^{\frac{n+r}{2}} = \Gamma\left(\frac{n+r}{2}\right) \left|\frac{1}{2}A\right|^{-\frac{n+r}{2}} {}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}BA^{-1}\right) $$

□

A.2.2 Proof of Proposition 4.2

Proof. To obtain the likelihood for the first observation, we have that k_1 is a gamma, Bauwens et al. (2000) gives the prior density for k_1 as:

$$|k_1|^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2} tr\big(k_1(1-\rho^2)\big)\Big)\frac{1}{c_0} \tag{A.2.1}$$

where $c_0 = \frac{\Gamma(\frac{n}{2})}{(\frac{1-\rho^2}{2})^{\frac{n}{2}}}$, is a constant and Γ is a gamma function. Let $V_1^{-1} = (1-\rho^2)$, thus, the likelihood for the first observation is as follows:

$$\begin{aligned} L(Y_1) &= \int L(Y_1 \mid k_1)\pi(k_1)dk_1 \\ &= \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}} k_1^{\frac{r}{2}} \exp\Big(-\frac{1}{2}\varepsilon_1^2 k_1\Big) k_1^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}(1-\rho^2)k_1\Big)\frac{1}{c_0}dk_1 \end{aligned} \tag{A.2.2}$$

The integral is with respect to k_1, so after rearranging and combining like terms we have;

$$L(Y_1) = \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}} k_1^{\frac{n+r-2}{2}} \exp\Big(-\frac{1}{2}(\varepsilon_1^2 + V_1^{-1})k_1\Big)\frac{1}{c_0}dk_1$$

where $k_1^{\frac{n+r-2}{2}} \exp(-\frac{1}{2}(\varepsilon_1^2 + V_1^{-1})k_1)$ is the kernel of a gamma with $n+r$ degrees of freedom. Let $\widetilde{V_2} = (\varepsilon_1^2 + V_1^{-1})^{-1}$, therefore, the density of $k_1|Y_1$ is:

$$\pi(k_1|Y_1) = k_1^{\frac{n+r-2}{2}} \exp\Big(-\frac{1}{2}k_1 \widetilde{V_2}^{-1}\Big)\frac{1}{\bar{c_0}} \tag{A.2.3}$$

with $\bar{c_0} = \frac{\Gamma(\frac{n+r}{2})}{(\frac{\widetilde{V_2}^{-1}}{2})^{\frac{n+r}{2}}}$. Thus, we have the likelihood as:

$$L(Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\Gamma\left(\frac{n+r}{2}\right)\left|\frac{\varepsilon_1^2 + V_1^{-1}}{2}\right|^{-\frac{n+r}{2}}\frac{1}{c_0}$$

Taking into account c_0 we can write the likelihood for $t=1$ as:

$$L(Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}} 2^{\frac{r}{2}} \frac{\Gamma(\frac{n+r}{2})}{\Gamma(\frac{n}{2})}\big|\varepsilon_1^2 + V_1^{-1}\big|^{-\frac{n+r}{2}} V_1^{-\frac{n}{2}}$$

Define $k_{1:2} = (k_1, k_2)$, then we have the likelihood for the second observation as:

$$L(Y_2|Y_1) = \int L(Y_2|k_{1:2}, Y_1)\pi(k_{1:2}|Y_1)dk_{1:2}$$

where $\pi(k_{1:2}|Y_1) = \pi(k_1|Y_1)\pi(k_2|k_1, Y_1)$.

The prior for k_t unconditionally is a gamma. However, $k_t|k_{t-1}$ is a non central

chisquared. (Muirhead, 2005, p.442) gives this non central chisquared density as follows:

$$\pi(k_t|k_{t-1}) = k_t^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_t\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_{t-1}k_t\right) \exp\left(-\frac{1}{2}\rho^2 k_{t-1}\right)\left(\Gamma\left(\frac{n}{2}\right)\right)^{-1}\frac{1}{c} \tag{A.2.4}$$

where ${}_0F_1$ is a hyper geometric function and $\rho^2 k_{t-1}$ is the non-centrality parameter and $c = 2^{\frac{n}{2}}$. Then we can write the likelihood for the second observation conditional on the first as :

$$L(Y_2|Y_1) = \int (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} k_2^{\frac{r}{2}} \exp\left(-\frac{1}{2}\varepsilon_2^2 k_2\right) \pi(k_{1:2}|Y_1) dk_{1:2} \tag{A.2.5}$$

We integrate first with respect to k_1. Define l_2 as representing all the elements in (A.2.4) that do not depend on k_1 as follows:

$$l_2 = \left(k_2^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_2\right)\right)^{-1}\left(\frac{1}{\Gamma(\frac{n}{2})}\right)^{-1}\left(\frac{1}{c}\right)^{-1} \tag{A.2.6}$$

Given that $\pi(k_2|k_1, Y_1) = \pi(k_2|k_1)$, and given (A.2.4) and (A.2.3), we can write $\pi(k_2|Y_1)$ as follows:

$$\pi(k_2|Y_1) = \int \pi(k_2|k_1, Y_1)\pi(k_1|Y_1)dk_1 =$$
$$\frac{1}{\bar{c_0}} \int k_1^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2}(\widetilde{V_2}^{-1}k_1)\right) \exp\left(-\frac{1}{2}(\rho^2 k_1)\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_1 k_2\right)\frac{1}{l_2}dk_1$$

where we have used the expression for $\pi(k_1|Y_1)$. We can write the above integral more compactly as:

$$\int \pi(k_2|k_1, Y_1)\pi(k_1|Y_1)dk_1 = \int \frac{1}{\bar{c_0}} k_1^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2}(\widetilde{V_2}^{-1} + \rho^2)k_1\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_1 k_2\right)\frac{1}{l_2}dk_1$$

Applying Lemma 5.1, we have the solution to this integral is as follows:

$$\pi(k_2|Y_1) = \int \pi(k_2|k_1, Y_1)\pi(k_1|Y_1)dk_1 = \\ \frac{1}{\bar{c_0}}\Gamma\left(\frac{n+r}{2}\right)\left|\frac{\widetilde{V_2}^{-1}+\rho^2}{2}\right|^{-\frac{n+r}{2}} {}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\right)\frac{1}{l_2} \tag{A.2.7}$$

Given (A.2.6) and (A.2.7), the distribution of $k_2|Y_1$ is a mixture of gamma's as follows:

$$\pi(k_2|Y_1) \propto k_2^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_2\right){}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\right) \tag{A.2.8}$$

The normalising constant for this hypergeometric function can be obtained in closed form, see Muirhead (2005, p.260). Thus, we have:

$$\int k_2^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_2\right){}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}k_2\delta_2\right)dk_2 = \Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}{}_2F_1\left(\frac{n+r}{2}, \frac{n}{2}; \frac{n}{2}; \delta_2\right) \tag{A.2.9}$$

where $\delta_2 = \rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}$. This ${}_2F_1\left(\frac{n+r}{2}, \frac{n}{2}; \frac{n}{2}; \delta_2\right)$ function has the same terms in the denominator and the numerator thus they cancel out and we have:

$${}_2F_1\left(\frac{n+r}{2}, \frac{n}{2}; \frac{n}{2}; \delta_2\right) = {}_1F_0\left(\frac{n+r}{2}; \delta_2\right) \tag{A.2.10}$$

This function simplifies to a known solution for $|\delta_2| < 1$, see Muirhead (2005, p.261) .

$${}_1F_0\left(\frac{n+r}{2}; \delta_2\right) = (1-\delta_2)^{-\frac{n+r}{2}} \tag{A.2.11}$$

The normalising constant becomes:

$$\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}{}_1F_0\left(\frac{n+r}{2}; \delta_2\right) = \Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}(1-\delta_2)^{-\frac{n+r}{2}}$$

Given this normalising constant, we have the density for $\pi(k_2|Y_1)$ from A.2.8 as follows:

$$\pi(k_2|Y_1) = \frac{1}{c_1}k_2^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_2\right){}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\right)$$

where $c_1 = \Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}(1-\delta_2)^{-\frac{n+r}{2}}$. Thus, the likelihood for the second observation is as follows:

$$L(Y_2|Y_1) = \int \pi(Y_2|k_2, Y_1)\pi(k_2|Y_1)dk_2$$
$$= \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}k_2^{\frac{n+r-2}{2}} \exp\bigg(-\frac{1}{2}(\varepsilon_2^2+1)k_2\bigg)\frac{1}{c_1}{}_1F_1\bigg(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}k_2\rho^2(\widetilde{V_2}^{-1}+\rho^2)^{-1}\bigg)dk_2$$

Using Muirhead (2005, p.261) and taking into account c_1, the likelihood for the second observation is:

$$L(Y_2|Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\frac{(\varepsilon_2^2+1)^{-\frac{n+r}{2}}}{(1-\delta_2)^{-\frac{n+r}{2}}}{}_2F_1\bigg(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; (\varepsilon_2^2+1)^{-1}\delta_2\bigg)$$

Thus we get a Gauss hypergeometric function which can be evaluated easily. Let $Z_2 = (\varepsilon_2^2+1)^{-1}\delta_2$. This series converges because $|Z_2| < 1$ (Jentschura et al., 1999). To accelerate the convergence of this series we apply the Euler transformation as in Srivastava & Choi (2012, p.67) and thus we get:

$${}_2F_1\bigg(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; Z_2\bigg) = (1-Z_2)^{-\frac{n+2r}{2}}{}_2F_1\bigg(-\frac{r}{2}, -\frac{r}{2}; \frac{n}{2}; Z_2\bigg) \tag{A.2.12}$$

Let $\hat{C}_2 = (1-Z_2)^{-\frac{n+2r}{2}}{}_2F_1\big(-\frac{r}{2}, -\frac{r}{2}; \frac{n}{2}; Z_2\big)$, then we can write the $L(Y_2|Y_1)$ as follows:

$$L(Y_2|Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\frac{(\varepsilon_2^2+1)^{-\frac{n+r}{2}}}{(1-\delta_2)^{-\frac{n+r}{2}}}\hat{C}_2$$

The density of k_t for the third observation is given by:

$$\pi(k_3|Y_2, Y_1) = \int \pi(k_3|k_2)\pi(k_2|Y_2, Y_1)dk_2$$

$\pi(k_2|Y_2, Y_1) \propto \pi(k_2|Y_1)L(Y_2|k_2, Y_1)$. The distribution for $\pi(k_2|Y_1)$ in (A.2.8) can be written as follows:

$$\pi(k_2|Y_1) \propto \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+2h_2-2}{2}} \exp\Big(-\frac{1}{2}k_2\Big)$$

where $\widetilde{C}_{2,h_2} = \frac{[(n+r)/2]_{h_2}}{[n/2]_{h_2}} \big(\frac{1}{2}\rho^2(\widetilde{V_2}^{-1} + \rho^2)^{-1}\big)^{h_2} \frac{1}{h_2!}$. Thus we have:

$$\pi(k_2|Y_2, Y_1) \propto \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+r+2h_2-2}{2}} \exp\Big(-\frac{1}{2}k_2(\varepsilon_2^2+1)\Big) \tag{A.2.13}$$

Given (A.2.4) and (A.2.13) we have:

$$\begin{aligned}\pi(k_3|Y_2, Y_1) \propto \int k_3^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_3\Big) {}_0F_1\Big(\frac{n}{2}; \frac{1}{4}\rho^2 k_2 k_3\Big) \exp\Big(-\frac{1}{2}\rho^2 k_2\Big) \\ \times \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+r+2h_2-2}{2}} \exp\Big(-\frac{1}{2}k_2(\varepsilon_2^2+1)\Big) \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}} dk_2\end{aligned}$$

which simplifies to:

$$\begin{aligned}\pi(k_3|Y_2, Y_1) \propto \int k_3^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_3\Big) {}_0F_1\Big(\frac{n}{2}; \frac{1}{4}\rho^2 k_2 k_3\Big) \exp\Big(-\frac{1}{2}(\varepsilon_2^2+1+\rho^2)k_2\Big) \\ \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+r+2h_2-2}{2}} \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}} dk_2\end{aligned}$$

Using Lemma 5.1 the density of $k_3|Y_2, Y_1$ is thus:

$$\begin{aligned}\pi(k_3|Y_2, Y_1) = \frac{1}{c_3} k_3^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_3\Big) \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\Big(\frac{n+r+2h_2}{2}\Big) \\ {}_1F_1\Big(\frac{n+r+2h_2}{2}; \frac{n}{2}; \frac{1}{2}k_3\rho^2 S_3\Big) (2S_3)^{\frac{n+r+2h_2}{2}} \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}\end{aligned} \tag{A.2.14}$$

where $S_3 = (\varepsilon_2^2 + 1 + \rho^2)^{-1}$ and c_3 is the normalising constant as in (A.2.9)as follows:

$$c_3 = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\Big(\frac{n+r+2h_2}{2}\Big) (2S_3)^{\frac{n+r+2h_2}{2}} {}_2F_1\Big(\frac{n+r+2h_2}{2}, \frac{n}{2}; \frac{n}{2}; \rho^2 S_3\Big)$$

Similar to (A.2.10) and (A.2.11), the hypergeometric function simplifies to get:

$$c_3 = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+r+2h_2}{2}\right)(2S_3)^{\frac{n+r+2h_2}{2}}(1-\rho^2 S_3)^{-\frac{n+r+2h_2}{2}}$$

Collecting terms dependent on h_2 we can write c_3 as

$$c_3 = \left(\sum_{h_2=0}^{\infty} \frac{[(n+r)/2]_{h_2}}{[n/2]_{h_2}}[(n+r)/2]_{h_2}\frac{\delta_3^{h_2}}{h_2!}\right)\Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$

where $\delta_3 = \left((1-\rho^2 S_3)^{-1} S_3 \rho^2 (\widetilde{V_2}^{-1} + \rho^2)^{-1}\right)$. This can be written as:

$$c_3 = {}_2F_1\left(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; \delta_3\right)\Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$

Using Euler's acceleration in (A.2.12) we have therefore:

$$c_3 = (1-\delta_3)^{-\frac{n+2r}{2}}\, {}_2F_1\left(-\frac{r}{2}, -\frac{r}{2}; \frac{n}{2}; \delta_3\right)\Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$

Therefore the likelihood for $t = 3$ is as follows:

$$L(Y_3|Y_2, Y_1) = \int \pi(Y_3|k_3, Y_2, Y_1)\pi(k_3|Y_2, Y_1)dk_3$$

Thus we have from (A.2.14)

$$L(Y_3|Y_2, Y_1) = \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}k_3^{\frac{n+r-2}{2}}\exp\left(-\frac{1}{2}k_3(\varepsilon_3^2+1)\right)\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+r+2h_2}{2}\right)$$
$${}_1F_1\left(\frac{n+r+2h_2}{2}; \frac{n}{2}; \frac{1}{2}k_3\rho^2 S_3\right)(2S_3)^{\frac{n+r+2h_2}{2}}\frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}dk_3$$

and we get:

$$L(Y_3|Y_2,Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+r+2h_2}{2}\right)(2S_3)^{\frac{n+r+2h_2}{2}}\Gamma\left(\frac{n+r}{2}\right)2^{\frac{n+r}{2}}$$
$$(\varepsilon_3^2+1)^{-\frac{n+r}{2}}{}_2F_1\left(\frac{n+r+2h_2}{2},\frac{n+r}{2};\frac{n}{2};(\varepsilon_3^2+1)^{-1}\rho^2S_3\right)\frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}$$

Applying the Euler transformation in (A.2.12) and letting $Z_3 = (\varepsilon_3^2+1)^{-1}\rho^2S_3$, we can define $\hat{C}_3 = (1-Z_3)^{-\frac{n+2r+2h_2}{2}}{}_2F_1\big(-\frac{r+2h_2}{2},-\frac{r}{2};\frac{n}{2};Z_3\big)$. Thus we have:

$$L(Y_3|Y_2,Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\frac{\Gamma\left(\frac{n+r+2h_2}{2}\right)}{(\varepsilon_3^2+1)^{\frac{n+r}{2}}}(2S_3)^{\frac{n+r+2h_2}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_3$$

The density for the fourth observation is given by:

$$\pi(k_4|Y_3,Y_2,Y_1) = \int \pi(k_4|k_3,Y_1,Y_2,Y_3)\pi(k_3|Y_3,Y_2,Y_1)dk_3 \tag{A.2.15}$$

with $\pi(K_3|Y_3,Y_2,Y_1) \propto \pi(K_3|Y_2,Y_1)L(Y_3|Y_2,Y_1)$. Let:

$$\widetilde{C}_{3,h_3} = \sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\left(\frac{n+r+2h_2}{2}\right)\frac{[(n+r)/2+h_2]_{h_3}}{[n/2]_{h_3}}\left(\frac{1}{2}\rho^2S_3\right)^{h_3}\frac{1}{h_3!}(2S_3)^{\frac{n+r+2h_2}{2}} \tag{A.2.16}$$

Then from (A.2.14) we have:

$$\pi(k_3|Y_2,Y_1) \propto \sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}k_3^{\frac{n+2h_3-2}{2}}\exp\left(-\frac{1}{2}k_3\right)$$

As before, when we include the third observation, the distribution of $k_3|Y_3,Y_2,Y_1$ is a mixture of gammas and can be written as follows:

$$\pi(k_3|Y_3,Y_2,Y_1) \propto \sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}k_3^{\frac{n+r+2h_3-2}{2}}\exp\left(-\frac{1}{2}k_3(\varepsilon_3^2+1)\right)$$

Let $\widetilde{V_4}^{-1} = (\varepsilon_3^2+1)$. Then, using (A.2.15), we have the distribution of $k_4|Y_3,Y_2,Y_1$ as follows:

$$\pi(k_4|Y_3,Y_2,Y_1) \propto \int k_4^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_4\Big)\,{}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_3 k_4\Big)\exp\Big(-\frac{1}{2}\rho^2 k_3\Big)\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}} \times \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} k_3^{\frac{n+r+2h_3-2}{2}} \exp\Big(-\frac{1}{2}k_3 \widetilde{V}_4^{-1}\Big)dk_3 \tag{A.2.17}$$

Taking this integral with respect to k_3 we get:

$$\pi(k_4|Y_3,Y_2,Y_1) \propto k_4^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_4\Big)\sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3}\,{}_1F_1\Big(\frac{n+r+2h_3}{2};\frac{n}{2};\frac{1}{2}\rho^2 k_4(\widetilde{V}_4^{-1}+\rho^2)^{-1}\Big) \Gamma\Big(\frac{n+r+2h_3}{2}\Big)(2S_4)^{\frac{n+r+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}$$

where $S_4 = (\widetilde{V}_4^{-1}+\rho^2)^{-1} = (\varepsilon_3^2+1+\rho^2)^{-1}$. Let c_4 be the normalising constant, that is:

$$c_4 = \int k_4^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_4\Big)\sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3}\,{}_1F_1\Big(\frac{n+r+2h_3}{2};\frac{n}{2};\frac{1}{2}\rho^2 k_4(\widetilde{V}_4^{-1}+\rho^2)^{-1}\Big) \Gamma\Big(\frac{n+r+2h_3}{2}\Big)(2S_4)^{\frac{n+r+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_4$$

Thus we get:

$$c_4 = \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3}\,{}_2F_1\Big(\frac{n+r+2h_3}{2},\frac{n}{2};\frac{n}{2};\frac{1}{2}\rho^2 S_4\Big)\Gamma\Big(\frac{n+r+2h_3}{2}\Big)(2S_4)^{\frac{n+r+2h_3}{2}}$$

Using (A.2.10) and (A.2.11), this simplifies to:

$$c_4 = \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3}(1-\rho^2 S_4)^{-\frac{n+r+2h_3}{2}}\Gamma\Big(\frac{n+r+2h_3}{2}\Big)(2S_4)^{\frac{n+r+2h_3}{2}}$$

Thus,

$$\pi(k_4|Y_3,Y_2,Y_1)=\frac{1}{c_4}k_4^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_4\Big)\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}{}_1F_1\Big(\frac{n+r+2h_3}{2};\frac{n}{2};\frac{1}{2}k_4(\widetilde{V}_4^{-1}+\rho^2)^{-1}\Big)$$
$$\Gamma\Big(\frac{n+r+2h_3}{2}\Big)(2S_4)^{\frac{n+r+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}$$

Therefore the likelihood for $t=4$ is as follows:

$$L(Y_4|Y_3,Y_2,Y_1)=\int\pi(Y_4|k_4,Y_3,Y_2,Y_1)\pi(k_4|Y_3,Y_2,Y_1)dk_4$$

Thus we have:

$$L(Y_4|Y_3,Y_2,Y_1)=\int(2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_4}k_4^{\frac{n+r-2}{2}}\exp\Big(-\frac{1}{2}k_4(\varepsilon_4^2+1)\Big)\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\Gamma\Big(\frac{n+r+2h_3}{2}\Big)$$
$${}_1F_1\Big(\frac{n+r+2h_3}{2};\frac{n}{2};\frac{1}{2}k_4\rho^2S_4\Big)(2S_4)^{\frac{n+r+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_4$$

This is similar to $t=3$ therefore we have:

$$L(Y_4|Y_3,Y_2,Y_1,\Sigma)=(2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_4}\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\frac{\Gamma\Big(\frac{n+r+2h_3}{2}\Big)}{(\varepsilon_4^2+1)^{\frac{n+r}{2}}}(2S_4)^{\frac{n+r+2h_3}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_4$$

and the likelihood for any t is:

$$L(Y_t|Y_{1:t-1})=(2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_t}\sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}\frac{\Gamma\Big(\frac{n+r+2h_{t-1}}{2}\Big)}{(\varepsilon_t^2+1)^{\frac{n+r}{2}}}(2S_t)^{\frac{n+r+2h_{t-1}}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_t$$

where:

$$\delta_t = \left((1-\rho^2 S_t)^{-1} S_t \rho^2 (\widetilde{V}_{t-1}^{-1} + \rho^2)^{-1} \right)$$

$$Z_t = (\varepsilon_t^2 + 1)^{-1} S_t \rho^2$$

$$\hat{C}_t = (1-Z_t)^{-\frac{n+2r+2h_{t-1}}{2}} {}_2F_1\left(-\frac{r+2h_{t-1}}{2}, -\frac{r}{2}; \frac{n}{2}; Z_t \right)$$

$$\widetilde{V}_t^{-1} = 1 + \varepsilon_{t-1}^2$$

$$S_t = (\varepsilon_{t-1}^2 + 1 + \rho^2)^{-1} = (\widetilde{V}_t^{-1} + \rho^2)^{-1}$$

$$c_t = \sum_{h_{t-1}=0}^{\infty} \widetilde{C}_{t-1,h_{t-1}} (1-\rho^2 S_t)^{-\frac{n+r+2h_{t-1}}{2}} \Gamma\left(\frac{n+r+2h_{t-1}}{2} \right) (2S_t)^{\frac{n+r+2h_{t-1}}{2}}$$

$$\widetilde{C}_{t-1,h_{t-1}} =$$

$$\sum_{h_{t-2}=0}^{\infty} \widetilde{C}_{t-2,h_{t-2}} \Gamma\left(\frac{n+r+2h_{t-2}}{2} \right) \frac{[(n+r)/2 + h_{t-2}]_{h_{t-1}}}{[n/2]_{h_{t-1}}} \left(\frac{1}{2}\rho^2 S_{t-1} \right)^{h_{t-1}} \frac{(2S_{t-1})^{\frac{n+r+2h_{t-2}}{2}}}{h_{t-1}!}$$

□

References

Abramowitz, M., Stegun, I. A., & Romer, R. H. (1988). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* American Association of Physics Teachers.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*(2), 177–190.

Banbura, M., Giannone, D., & Reichlin, L. (n.d.). *Bayesian vars with large panels* (Discussion Paper No. DP6326). London: CEPR Discussion Paper No. DP6326.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of applied Econometrics*, *25*(1), 71–92.

Bank of England. (2022). *Monetary policy report may 2022.* `https://www.bankofengland.co.uk/monetary-policy-report/2022/may-2022`. (Last checked on Apr 08, 2023)

Bank of England. (2023). *Quantitative easing.* `https://www.bankofengland.co.uk/monetary-policy/quantitative-easing`. (Last checked on Jul 04, 2023)

Bauwens, L., Lubrano, M., & Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models*. OUP Oxford.

Bernardo, J. M. (1979). Expected information as expected utility. *the Annals of Statistics*, *7*(3), 686–690.

Bishop, C. M. (2007). *Pattern recognition and machine learning (information science and statistics)*. Springer New York.

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American statistical association business and economic statistics section*.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542–547.

Brian, D., LInda, K., & Paul, W. (2007). *The experience of foreign central banks with published forecasts.* `https://www.federalreserve.gov/monetarypolicy/files/FOMC20070103memo01.pdf`. (Last checked on Apr 08, 2023)

Britton, E., Fisher, P., & Whitley, J. (n.d.). The inflation report projections: understanding the fan chart. *Bank of England Quarterly Bulletin*, *38*(1998), 30–37.

Canova, F., Ciccarelli, M., & Ortega, E. (2007). Similarities and convergence in g-7 cycles. *Journal of Monetary economics*, *54*(3), 850–878.

Carriero, A., Clark, T. E., & Marcellino, M. (2016). Common drifting volatility in large bayesian vars. *Journal of Business & Economic Statistics*, *34*(3), 375–390.

Carriero, A., Clark, T. E., & Marcellino, M. (2018). Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics*, *100*(5), 799–815.

Chan, J. C. (2013). Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, *176*(2), 162–172.

Chan, J. C. (2020). Large bayesian vars: A flexible kronecker error covariance structure. *Journal of Business & Economic Statistics*, *38*(1), 68–79.

Chan, J. C. (2023). Comparing stochastic volatility specifications for large bayesian vars. *Journal of Econometrics*, *235*(2), 1419–1446.

Chan, J. C., Doucet, A., León-González, R., & Strachan, R. W. (2020, sep). *Multivariate stochastic volatility with co-heteroscedasticity* (Discussion Paper). GRIPS Discussion Paper No. 20-09.

Chan, J. C., & Grant, A. L. (2016a). Modeling energy price dynamics: Garch versus stochastic volatility. *Energy Economics*, *54*(C), 182–189.

Chan, J. C., & Grant, A. L. (2016b). On the observed-data deviance information criterion for volatility modeling. *Journal of Financial Econometrics*, *14*(4), 772–802.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the american statistical association*, *90*(432), 1313–1321.

Chib, S., Nardari, F., & Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, *134*(2), 341–371.

Chiu, C.-W. J., Mumtaz, H., & Pinter, G. (2017). Forecasting with var models: Fat tails and stochastic volatility. *International Journal of Forecasting*, *33*(4), 1124–1143.

Clemen, R. T., Murphy, A. H., & Winkler, R. L. (1995). Screening probability forecasts: contrasts between choosing and combining. *International Journal of Forecasting*, *11*(1), 133–145.

Cogley, T., & Sargent, T. J. (2001). Evolving post-world war ii us inflation dynamics. *NBER macroeconomics annual*, *16*, 331–373.

Creal, D. D. (2017). A class of non-gaussian state space models with exact likelihood inference. *Journal of Business & Economic Statistics*, *35*(4), 585–597.

Cross, J., & Poon, A. (2016). Forecasting structural change and fat-tailed events in australian macroeconomic variables. *Economic Modelling*, *58*, 34–51.

David, Reifschneider and Peter, Tulip. (2017). *Fan charts.* `https://www.rba.gov.au/publications/rdp/2017/2017-01/fan-charts.html`. (Last checked on Apr 08, 2023)

DeGroot, M. H., & Fienberg, S. E. (1981). *Assessing probability assessors: Calibration and refinement.* (Tech. Rep.). Carnegie-Mellon University Pittsburgh PA Department of Statistics.

Diebold, F. X., & Lopez, J. A. (1996). 8 forecast evaluation and combination. *Handbook of statistics*, *14*, 241–268.

Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, *3*(1), 1–100.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, *82*(4), 540–554.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* CRC press.

Geweke, J. (2004). Getting it right: Joint distribution tests of posterior simulators. *Journal of the American Statistical Association*, *99*(467), 799–804.

Geweke, J. (2005). *Contemporary bayesian econometrics and statistics.* John Wiley & Sons.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130–141.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *14*(1), 107–114.

Götz, T. B., & Hauzenberger, K. (2021). Large mixed-frequency vars with a parsimonious time-varying parameter structure. *The Econometrics Journal*, *24*(3), 442–461.

Gouriéroux, C., & Jasiak, J. (2006). Autoregressive gamma processes. *Journal of forecasting*, *25*(2), 129–152.

Groen, J. J., Kapetanios, G., & Price, S. (2009). A real time evaluation of bank of england forecasts of inflation and growth. *International Journal of Forecasting*, *25*(1), 74–80.

Hartwig, B. (2022). Bayesian vars and prior calibration in times of covid-19. *Studies in Nonlinear Dynamics & Econometrics*, *Just Accepted*.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hou, C., Nguyen, B., & Zhang, B. (2023). Real-time forecasting of the australian macroeconomy using flexible bayesian vars. *Journal of Forecasting*, *42*(2), 418–451.

Jacquier, E., Polson, N. G., & Rossi. (1994). Bayesian analysis of stochastic volatility models(with discussion). *journal of business & economic statistics*, *12*(1), 371–389.

Jentschura, U. D., Mohr, P. J., Soff, G., & Weniger, E. J. (1999). Convergence acceleration via combined nonlinear-condensation transformations. *Computer physics communications*, *116*(1), 28–54.

Jun Saito. (2022). *Foreign exchange market intervention and monetary policy.* `https://www.jcer.or.jp/english/foreign-exchange-market-intervention-and-monetary-policy`. (Last checked on Jul 04, 2023)

Kadiyala, K. R., & Karlsson, S. (1997). Numerical methods for estimation and inference in bayesian var-models. *Journal of Applied Econometrics*, *12*(2), 99–132.

Karlsson, S. (2013). Forecasting with bayesian vector autoregression. *Handbook of economic forecasting*, *2*, 791–897.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, *65*(3), 361–393.

Konishi, S., & Kitagawa, G. (2008). *Information criteria and statistical modeling*. Springer Science & Business Media.

Koop, G. M. (2013). Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics*, *28*(2), 177–203.

Kroese, D. P., Taimre, T., & Botev, Z. I. (2013). *Handbook of monte carlo methods*. John Wiley & Sons.

Lehmann, E. L., & Romano, J. P. (1986). *Testing statistical hypotheses* (Vol. 3). Springer.

Lemos, M. C., Finan, T. J., Fox, R. W., Nelson, D. R., & Tucker, J. (2002). The use of seasonal climate forecasting in policymaking: lessons from northeast brazil. *Climatic Change*, *55*, 479–507.

Leon-Gonzalez, R. (2018). Efficient bayesian inference in generalized inverse gamma processes for stochastic volatility. *Econometric Reviews*, *38*(8), 899–920.

Leon-Gonzalez, R., et al. (2021). Forecasting macroeconomic variables in emerging economies. *Journal of Asian Economics*, *77*, 101–403.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Madan, D. B., & Seneta, E. (1990). The variance gamma model for share market returns. *Journal of business*, 511–524.

Mishkin, F. S. (2004). *Can central bank transparency go too far?* (Working Paper No. w10829). USA.

Mitchell, J., & Hall, S. G. (2005). Evaluating, comparing and combining density forecasts using the klic with an application to the bank of england and niesr 'fan'charts of inflation. *Oxford bulletin of economics and statistics*, *67*, 995–1033.

Moral, P. (2004). *Feynman-kac formulae: genealogical and interacting particle systems with applications*. Springer.

Muirhead, R. J. (2005). *Aspects of multivariate statistical theory* (Vol. 197). John Wiley & Sons.

Mumtaz, H. (2016). The evolving transmission of uncertainty shocks in the united kingdom. *Econometrics*, *4*(1), 1-18.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, *12*(2), 758–765.

Omori, Y., Chib, S., Shephard, N., & Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, *140*(2), 425–449.

Par, O. (2006). *Incorporating judgements in fan charts.* `https://www.federalreserve.gov/econres/feds/incorporating-judgement-in-fan-charts.html`. (Last checked on Apr 08, 2023)

Poon, A. (2018). Assessing the synchronicity and nature of australian state business cycles. *Economic Record*, *94*(307), 372–390.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, *81*(10.5555), 26853.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica sinica*, *7*(2), 221–242.

Shephard, N. (1994). Local scale models: State space alternative to integrated garch processes. *Journal of Econometrics*, *60*(1-2), 181–202.

Shephard, N., & Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, *84*(3), 653–667.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika*, *68*(1), 45–54.

Shibata, R. (1983). Asymptotic mean efficiency of a selection of regression variables. *Annals of the Institute of Statistical Mathematics*, *35*, 415–423.

Shuford Jr, E. H., Albert, A., & Edward Massengill, H. (1966). Admissible probability measurement procedures. *Psychometrika*, *31*(2), 125–145.

Sims, C., & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, *39*(4), 949–968.

Sims, C., & Zha, T. (2006). Were there regime switches in macroeconomic policy? *American Economic Review*, *96*(1), 54–81.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, *48*(1), 1–48.

Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica: Journal of the Econometric Society*, *5*(2), 105–146.

South African Reserve Bank. (2021). *Reserves management and foreign exchange operations.* `https://www.resbank.co.za/en/home/what-we-do/financial-markets/foreign-exchange`. (Last checked on Jul 03, 2023)

Srivastava, H. M., & Choi, J. (2012). *Zeta and q-zeta functions and associated series and integrals*. Elsevier.

Statista Research Department. (2022). *Quantitative easing by the bank of england.* `https://www.statista.com/statistics/1105570/value-of-quantitative-easing-by-the-bank-of-england-in-the-united-kingdom`. (Last checked on Jul 04, 2023)

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 44–47.

Sundararajan, R. R., & Barreto-Souza, W. (2023). Student-t stochastic volatility model with composite likelihood em-algorithm. *Journal of Time Series Analysis*, *44*(1), 125–147.

Svensson, L. E. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European economic review*, *41*(6), 1111–1146.

Theil, H. (1958). *Economic policy and forecasts.* North-Holland: Amsterdam, The Netherlands.

Tinbergen, J. (n.d.). Centralization and decentralization in economic policy. *The Economic Journal*, *66*(261), 133–134.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica: Journal of the Econometric Society*, *65*(1), 59–73.

Wang, J. J., Chan, J. S., & Choy, S. B. (2011). Stochastic volatility models with leverage and heavy-tailed distributions: A bayesian approach using scale mixtures. *Computational Statistics & Data Analysis*, *55*(1), 852–862.

West, M., & Harrison, J. (2006). *Bayesian forecasting and dynamic models.* Springer Science & Business Media.

Wieland, V., & Wolters, M. (2013). Forecasting and policy making. In *Handbook of economic forecasting* (Vol. 2, pp. 239–325). Elsevier.

Winkler, R. L. (1969). Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association*, *64*(327), 1073–1078.

Winkler, R. L., & Murphy, A. H. (1968). "good" probability assessors. *Journal of Applied Meteorology and Climatology*, *7*(5), 751–758.

World Bank. (2022). *World bank country and lending groups.* `https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups`. (Last checked on Jul 02, 2023)

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, *127*(2), 165–178.

List of Tables

List of Figures