Bayesian Inference of State Space Models with Flexible Covariance Matrix Rank: Applications for Inflation Modeling

Luis Uzeda

A thesis submitted for the degree of
Doctor of Philosopy at
The Australian National University

February 2017

Except where otherwise indicated, this thesis is my own original work.

Luis Uzeda
16 February 2017

I dedicate this dissertation to my partner and best friend Elodie, to my daughter Laís and to all my family in Brazil for their endless support.

Acknowledgements

First I would like to express my gratitude to my supervisory panel members: Professor Rodney Strachan, Professor Joshua Chan and Dr. Timothy Kam. Professor Rodney Strachan's mentorship precedes my PhD journey and I will be forever grateful for his support over the past decade. Professor Joshua Chan's encouragement, patience and support over the past five years has been paramount. I could not possibly thank him enough for the uncountable amount of times I dropped by his office asking if he would have a minute for a quick question which more often than not turned into many minutes of long questions. Dr Kam kindly joined my panel as the last member. His advice was very constructive and useful. Especially during my presentations at the RSE macro study group.

I am also greatly indebted to the research department at the Reserve Bank of New Zealand. I thoroughly enjoyed the time I spent there during my dissertation internship in 2016. My job market paper has benefited tremendously from the feedback provided by everyone in the research team.

I would like to thank the ANU Research School of Economics for funding through the ANU Research Scholarship and Teaching Fellowship Programs as well as thank my friends, both inside and outside ANU, who spent their time studying and discussing research and non-research related issues with me over the past five years.

This acknowledgement would not be complete without a mention to my family. I would like to thank my family in Brazil for their unwavering support and love through all the different periods of my life. Last but definitely not least, I would like to thank my partner Elodie and my daughter Laís. None of this would have been possible without your love, understanding and patience throughout this journey. Je vous aime!

Abstract

After the introductory chapter, this thesis comprises two further chapters. The main chapters in this dissertation, i.e., chapters 2 and 3 are presented in essay format, each with an independent introduction and conclusion. The contents of these individual chapters are outlined below.

Chapter 2 studies the forecasting implications of specifying unobserved components (UC) models with different state correlation structures. While implications for signal extraction from specifying UC models with correlated or orthogonal innovations have been well-investigated, out-of-sample implications are less well understood. This paper attempts to address this gap in light of the recent resurgence of studies adopting UC models for forecasting purposes. Four correlation structures for errors are entertained: orthogonal, correlated, perfectly correlated innovations as well as a novel approach which combines features from two contrasting cases, namely, orthogonal and perfectly correlated innovations. Parameter space restrictions associated with different correlation structures and their connection with forecasting are discussed within a Bayesian framework. As perfectly correlated innovations reduce the covariance matrix rank, a Markov Chain Monte Carlo sampler which builds upon properties of Toeplitz matrices and recent advances in precision-based algorithms is developed. Our results for several measures of U.S. inflation indicate that the correlation structure between state variables has important implications for forecasting performance as well as estimates of trend inflation.

Chapter 3 develops an econometric framework to investigate the contribution of monetary policy to the evolution of U.S. trend inflation. We combine two modeling approaches – measuring trend inflation using an unobserved components model and estimation of monetary policy rules with drifting coefficients – to investigate interdependence between policy rule parameters and trend inflation. We employ identification strategies of the policy shock to trend inflation which highlight particular changes in the conduct of systematic monetary policy and overidentify a state space model for inflation and the policy rate. An efficient Markov Chain Monte Carlo algorithm using precision-based methods is proposed for static and dynamic selection of policy drivers behind trend inflation. Our empirical analysis indicates three main results: (1) the influence of monetary policy on trend inflation increased during the Great Moderation relative to the Great Inflation period; (2) non-policy shocks, however, accounted for between $50\% - 70\%$ of the variation in trend inflation during

each of these episodes; (3) monetary policy's contribution to stabilize trend inflation around the early 1980s reflects a weaker reaction to output gap changes accompanied by a stronger emphasis on inflation gap dynamics and inflation target adjustments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

State space models are powerful tools for characterizing a range of time series features of interest such as trends as well as cyclical and seasonal patterns. This flexibility, however, can lead to model overfitting. State variables —used to capture unobserved features of interest— are, after all, parameters which can vary over time. Therefore, the inference problem in a state space framework can be broadly characterized as demanding the data to be informative on a number of parameters which can far exceed the amount of observations which are available.

The brief discussion above is intended to motivate an existing (and well known) trade-off between modeling flexibility and inference performance. In this thesis, we study and develop Bayesian state space inference techniques which address issues between preserving specification flexibility and inferential accuracy. The overarching methodological approach in this dissertation hinges on adopting a flexible rank structure of the covariance matrix for state space models. More specifically, we accommodate covariance matrices with flexible rank by allowing for a range of correlation structures between innovations driving the states. Covariance matrices become rank deficient as a result of introducing perfect correlation between some (or among all) states.

Chapter 2 studies the forecasting implications of specifying UC models with different correlation structures. In keeping with the recent literature, we adopt a Bayesian approach to conduct the forecasting exercise. The Bayesian framework provides a useful strategy to investigate the relationship between correlation and the forecasting function. To be clear, forecasting within a Bayesian setting requires integration of the predictive distribution over the parameter space of a model. Moreover, by specifying an UC model under its reduced form autoregressive moving average representation, helps visualizing how assumptions about the correlation between state innovations lead to different restrictions over the parameter space for the . Since the parameter space of a model is the support of the forecasting function, it becomes an empirical question whether or not one should specify UC models which deviate

from the common orthogonal innovations assumption.

The last point above is studied at length in this chapter. In particular, four correlation structures for errors are employed: orthogonal, correlated, perfectly correlated innovations as well as a novel approach which combines features from two contrasting cases, namely, orthogonal and perfectly correlated innovations. As perfectly correlated innovations reduce the covariance matrix rank, a Markov Chain Monte Carlo sampler which builds upon properties of Toeplitz matrices and recent advances in precision-based algorithms is developed. The algorithm is generalized to accommodate estimation of UC models exhibiting both full and reduced rank covariance matrices. A substantial application for inflation forecasting illustrates the importance of correlation within UC models for out-of-sample analysis. In summary, three contributions emerge from this chapter: (1) a comprehensive discussion on Bayesian forecasting implications that stem from specifying UC models with different correlation structures; (2) a new class of models that bridges the existing orthogonal and perfectly correlated innovations UC models; (3) a novel algorithm for fast disturbance smoothing based on precision sampling techniques.

Chapter 3 develops an econometric framework to investigate the contribution of monetary policy to the dynamics of U.S. trend inflation. We combine two modeling approaches – measuring trend inflation using an unobserved components model and estimation of monetary policy rules with drifting coefficients – to investigate interdependence between policy rule parameters and trend inflation. We seek identification strategies of the policy shock to trend inflation which highlight a particular change in the conduct of systematic monetary policy and overidentify a state space model for inflation and the policy rate. An efficient Markov Chain Monte Carlo algorithm which builds upon recent advances in precision-based methods is then proposed for static and dynamic selection of policy drivers behind trend inflation. Our analysis indicates three main results: (1) the influence of monetary policy on trend inflation increased during the Great Moderation relative to the Great Inflation period; (2) non-policy shocks, however, accounted for around 69% and 50% of the variation in trend inflation during each of these episodes; (3) monetary policy contribution to stabilize trend inflation around the early 1980s reflects a weaker reaction to output gap changes accompanied by a stronger emphasis on inflation gap dynamics and inflation target adjustment shocks.

Finally, Chapter 4 offers concluding remarks.

Chapter 2

State Correlation and Forecasting: A Bayesian Approach Using Unobserved Components Models

2.1 Introduction

In this chapter we study various correlation structures for innovations within unobserved components (UC) models. The implications for Bayesian forecasting are studied from a theoretical and empirical standpoint. Four main contributions emerge from our analysis: (1) a comprehensive discussion of the forecasting implications that stem from specifying UC models with different correlation structures within a Bayesian framework; (2) a new class of UC models that bridges the existing orthogonal and perfectly correlated innovations schemes; (3) a new algorithm for fast estimation of state space models based on precision sampling techniques; and (4) a substantial inflation forecasting application incorporating various specifications of UC models.

Unobserved components models provide a flexible and yet parsimonious framework which has been widely employed in empirical macroeconomics over the years.[1] When estimating such models, however, one is typically confronted with the issue of formulating the correlation structure between innovations driving different components (states). Depending on the subject matter, economic theory, statistical properties or a combination of both can be used to provide guidance on suitable modelling strategies of the covariance matrix. In particular, there is a wealth of studies discussing different approaches to model correlation and their effects on estimated outputs (i.e. implications to a signal extraction problem). For example, Harvey and

[1] We cannot possibly do justice to the literature here. We point the reader to Harvey [1985], Watson [1986], Clark [1987], Morley et al. [2003], Proietti [2006], Stock and Watson [2007], Perron and Wada [2009] and Luo and Startz [2014] for an overview of the literature.

Koopman [2000] and Proietti [2006] investigate the differences in parametric filtering that arise in terms of how observations are weighted when adopting UC models with orthogonal and correlated innovations.[2] Others, such as Morley et al. [2003], Oh and Zivot [2006], Oh et al. [2008] and Iwata and Li [2015] study different correlation structures to reconcile discrepancies in business cycle measures generated by UC models and the ones based on the Beveridge-Nelson decomposition (see Beveridge and Nelson [1981]). Similarly, Dungey et al. [2015] adopt a correlated innovations UC model framework to suggest identification strategies of permanent and transitory shocks to trend and cyclical components of US real GDP.

The brief description above is intended to highlight that many authors have studied what can be broadly interpreted as *in-sample* implications of different correlation structures within UC models. However, a corresponding comparative study in terms of *out-of-sample* implications remains, to the best of our knowledge, uninvestigated to date. As such, the analysis provided in this chapter contributes to fill this gap. Moreover, the recent resurgence of papers adopting UC models for inflation forecasting purposes as in Stock and Watson [2007], Chan, Koop and Potter (2013), Stella and Stock (2013), Garnier, Mertens and Nelson (2013), Chan [2013] and Clark and Doh (2014) strengthens the case for out-of-sample evaluation of the correlation structure within such models.

Notably, modern approaches to forecast using UC models, as in the studies mentioned above, typically exhibit three features: (1) using Bayesian, or more precisely, Markov Chain Monte Carlo (MCMC) techniques to conduct estimation; (2) orthogonal innovations; and (3) introducing stochastic volatility a la, e.g., Kim et al. [1998] to model changes in the conditional variance of innovations over time. In this chapter we keep point 1, extend point 2 and leave point 3 for future research. To be clear, in our empirical exercise, we decide to leave out stochastic volatility not because we think it is unimportant for forecasting, but in order to make the out-of-sample assessment of each correlation structure as free as possible from other modeling features that can also impart changes to forecasting performance. Once we are confident about potential forecasting benefits of UC models which deviate from the popular orthogonal innovations framework, a natural extension to our study is to incorporate stochastic volatility and conduct the forecasting exercise to a broader range of models.

It is important to recognize that when generating forecasts which incorporate parameter uncertainty, as in the case of Bayesian estimation, the correlation structure

[2]Harvey and Koopman [2000] show that orthogonal innovations imply two-sided filters (or smoothers) with symmetric weights in the middle of a series. Such symmetry is argued by the authors to be an attractive feature of UC models with orthogonal components.

has a direct connection with the forecasting function which does not arise naturally if adopting other approaches such as maximum likelihood based forecasts. Within a Bayesian setting, construction of the predictive density entails integration of the likelihood function over the parameter space of a model. As discussed in Section 2.3, depending on how one models the correlation between innovations, restrictions on the parameter space associated with a particular UC representation can be imposed or relaxed and, by the same token, point and density forecasts can be affected. In particular, the usual orthogonal innovations approach often imposes strong parameter space restrictions when compared to their correlated counterparts, as pointed out in, e.g, Harvey and Koopman [2000], Morley et al. [2003], Ord et al. [2005] and Oh et al. [2008]. Obviously, whether a more or less restricted parameter space is desirable for forecasting performance is an empirical question which we address in this chapter.

Our empirical evaluation is built around four correlation structures, namely, innovations (or, equivalently, states) are allowed to be orthogonal, correlated, perfectly correlated as well as a new approach which combines aspects from two contrasting correlation structures. In particular, we construct UC models which specify two latent components driven by the same stochastic process (i.e. perfect correlation) while the dynamics of a third component are governed by an orthogonal innovation. As a result, we propose a new class of UC model which bridges the usual orthogonal innovations approach (e.g. Harvey [1985], Clark (1987), Stock and Watson [2007]) and the single source of error (SSOE) representation of state space models advocated by, for example, Snyder [1985], Ord et al. [1997] and Chatfield et al. [2001]. We refer to such a class of models as reduced source of error (RSOE) to distinguish it from its SSOE and multiple source of error (MSOE) counterparts. When modeling inflation, for example, RSOE models can combine an orthogonal trend inflation component, as commonly adopted, with a flexible and innovations-parsimonious representation of transitory inflation dynamics. In addition, as we show later, the RSOE scheme can also represent a compromise between SSOE and MSOE variants in terms of parameters space restrictions.

To evaluate forecast performance across different correlation structures, a substantive forecasting exercise accounting for eleven UC models is conducted. In keeping with other studies using UC models for forecasting purposes (e.g. Stock and Watson [2007], Chan [2013], Chan [2015], Clark and Doh [2014] and Garnier et al. [2015]), we use several quarterly inflation measures in our empirical application. Forecasting performance is assessed in terms of point and density forecast accuracy. We find that the choice of correlation structure between state variables has appreciable implications to forecasting performance. Allowing for correlation between innovations leads to statistically significant improvements in both point and density forecast estimates

at various forecasting horizons relative to orthogonal innovations. In addition, even though the focus here is on forecasting, we show that trend inflation measures can be sensitive to different correlation structures. In particular, RSOE models generate smoother measures of trend inflation, which is often perceived as a desirable feature for policy analysis. To gauge the statistical significance of the differences in forecasting performance, we follow other authors (see, e.g., Bauwens et al. [2014], Clark and Ravazzolo [2014], Clark and Doh [2014] and Garnier et al. [2015]) and report t-test results for the Diebold and Mariano (1995) test. Allowing for perfectly correlated innovations (as in the SSOE and RSOE cases) produces covariance matrices with reduced rank. To accommodate rank reduction, we develop an MCMC sampler which builds upon properties of Toeplitz matrices and extend previous work on precision-based algorithms for state space models in Chan and Jeliazkov [2009]. In particular, we propose a new (precision-based) disturbance smoothing algorithm which adds to the existing (Kalman-filter based) ones of DeJong and Shephard (1995) and Durbin and Koopman (2002). As shown in McCausland et al. [2011] precision-based samplers are more efficient than the traditional Kalman filter-based approach for state simulation.

The structure of the rest of this chapter is as follows: Section 2.2 presents all UC models entertained in this chapter. In Section 2.3 we discuss how changes in state correlation can affect the forecasting distribution within a Bayesian estimation framework. Section 2.4 deals with the issues of carrying out MCMC estimation of UC models with a reduced rank covariance matrix. Section 2.5 develops an efficient and general posterior simulator to estimate UC models with both full and reduced rank covariance matrices. Out-of-sample forecast evaluation based on various correlation structures is presented in Section 2.6. Section 2.7 concludes and presents directions for future research.

2.2 The Models

We begin by presenting the UC models adopted in our empirical application. In particular, models can be organized in two categories, namely, UC models suited for I(1) and I(2) univariate processes (I(1)-UC and I(2)-UC hereafter). Since our focus is on inflation, I(1)-UC models are fit to the first difference in log price level measures while I(2)-UC models address inflation dynamics by directly modelling movements in the log price level (e.g. log CPI and log real GDP implicit price deflator).[3] The I(2)-UC models lead to a larger number of innovations and, consequently, a wider range of correlation structures relative to I(1)-UC models are explored.[4]

2.2.1 I(1)-UC models

Let y_t denote an univariate I(1) process such that

$$y_t = \tau_t + c_t, \tag{2.1a}$$

$$\tau_t = \tau_{t-1} + \eta_t, \tag{2.1b}$$

$$\phi(L)c_t = \varepsilon_t, \tag{2.1c}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta \\ \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 \end{bmatrix} \right). \tag{2.1d}$$

Therefore, I(1)-UC models describe y_t as the sum of two latent components, each of which being responsible for different type of dynamics. Specifically, when y_t denotes an inflation measure, τ_t is commonly referred as *trend inflation*, which accords well with the Beveridge-Nelson characterization of what consists long-run dynamics in macroeconomic aggregates (see Beveridge and Nelson [1981]). Transitory deviations about τ_t are captured by an ergodic autoregressive process, c_t. In keeping with previous studies (e.g. Kang et al. [2009] and Garnier et al. [2015]) such transient dynamics provide a measure for the *inflation gap*. To allow for persistent movements in c_t, a p^{th}-order autoregressive lag polynomial, $\phi(L) = (1 - \phi_1 L^1 - \phi_2 L^2 - ... - \phi_p L^p)$, is introduced with roots of $\phi(x) = 0$ lying outside the unit circle. In particular, we consider two cases: $p = 0$ and $p = 2$, hence modeling c_t as an AR(0) or AR(2) process, respectively. Our choice of p is motivated by previous studies.

[3] Data description is deferred to Section 2.6.1.

[4] Admittedly, the order of integration of inflation is a debatable issue. For example, depending on the sample period, Dickey-Fuller type of tests suggest one or no unit-root for all inflation measures. Given the focus here is on out-of-sample performance, we take an agnostic view on pre-testing procedures and produce forecasts based on models which assume inflation and price level series are I(1) and I(2) respectively. We stress, however, that Dickey-Fuller test results based on the full sample used in our forecasting exercise detected one unit-root and two unit roots for inflation and log price level measures, respectively.

When $p = 0$ (i.e. $\phi(L) = 1$) the framework above describes a simple random walk plus noise, or local level, model. Such class of model has been adopted recently by numerous inflation forecasting studies (e.g. Stock and Watson [2007], Chan [2013], Clark and Doh [2014]). In contrast, when c_t is an AR(2) process (as in, e.g., Kang et al. [2009] and Garnier et al. [2015]) persistence in c_t might reflect, for example, the role of nominal price rigidities in slowing down price adjustments about τ_t, as postulated by New Keneysian macroeconomic models (e.g. Christiano et al. [2005] and Smets and Wouters [2007]). In addition, Morley et al. [2003] show that specifying c_t as an AR(2) process enables just-identification of $\rho_{\varepsilon\eta}$.[5] As such, $\rho_{\varepsilon\eta}$ can be inferred using sample information rather than being fixed according to some arbitrary identification strategy. This contrasts with the local level model, where different values of $\rho_{\varepsilon\eta}$ can lead to equivalent evaluations of the likelihood function (see e.g. Harvey [1989], chapter 2 and Morley et al. [2003]), hence making estimation of this parameter more difficult.

A common strategy to address identification of $\rho_{\varepsilon\eta}$ is to set $\rho_{\varepsilon\eta} = 0$, such that we treat ε_t and η_t as being orthogonal (e.g. Stock and Watson [2007], Chan [2013] and Clark and Doh [2014]). Other authors, such as Ord et al. [1997], Snyder et al. [2001] and Chatfield et al. [2001] set $\rho_{\varepsilon\eta} = \pm 1$, such that state innovations are perfectly correlated. These contrasting identification strategies, however, affect the construction of predictive densities. We discuss this issue more carefully in Section 2.3 as it constitutes an important motivation to our empirical exercise.

The three structures entertained in this chapter to model the covariance matrix, Ω, associated with I(1)-UC models are:

$$\Omega = \left\{ \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta \\ \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 \end{bmatrix}, \quad \begin{bmatrix} \sigma_\varepsilon^2 & \pm\sigma_\varepsilon\sigma_\eta \\ \pm\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 \end{bmatrix}, \quad \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \right\}.$$

The first denotes the unrestricted case where $\rho_{\eta\varepsilon}$ is estimated. As discussed above, this covariance structure only applies to the case where c_t is set as an AR(2) process. The second and third structures describe the perfectly correlated and uncorrelated innovations cases, respectively.

Now, let κ_τ denote a loading parameter and MNZ be the short notation for the UC model in Morley, Nelson and Zivot (2003) which sets c_t as an AR(2) process. Combining the three covariance structures above with $p = 0$ and $p = 2$, accordingly, gives rise to five I(1)-UC models with the following state equations:

- **Local Level-SSOE** ($\rho_{\varepsilon\eta} = \pm 1$, and $p = 0$): $\tau_t = \tau_{t-1} + \kappa_\tau\varepsilon_t$, $c_t = \varepsilon_t$,

[5] To be precise, by just-identification of $\rho_{\varepsilon\eta}$ (or any other parameter) we mean that the likelihood contribution is not invariant to different values of $\rho_{\varepsilon\eta}$.

- **Local Level-MSOE** ($\rho_{\varepsilon\eta} = 0$, and $p = 0$): $\tau_t = \tau_{t-1} + \eta_t$, $c_t = \varepsilon_t$,
- **MNZ-SSOE** ($\rho_{\varepsilon\eta} = \pm 1$, and $p = 2$): $\tau_t = \tau_{t-1} + \kappa_\tau \varepsilon_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$,
- **MNZ-MSOE(UR)** ($\rho_{\varepsilon\eta} \neq 0$, and $p = 2$): $\tau_t = \tau_{t-1} + \eta_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$,
- **MNZ-MSOE** ($\rho_{\varepsilon\eta} = 0$, and $p = 2$): $\tau_t = \tau_{t-1} + \eta_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$.

Since the UC model in MNZ can accommodate two variants of MSOE schemes, namely, when $\rho_{\varepsilon\eta} = 0$ and $\rho_{\varepsilon\eta}$ is unrestricted, we refer to each case as MNZ-MSOE and MNZ-MSOE(UR), respectively. When innovations are allowed to be correlated (i.e. SSOE and MSOE(UR)), the loading parameter, κ_τ, governs the correlation sign between states as well as the magnitude of the effect the common innovation, ε_t, has on τ_t. In particular, when $\rho_{\varepsilon\eta}$ is unrestricted we follow Luo and Startz [2014] and specify innovations to τ_t as $\eta_t = \eta_t^* + \kappa_\tau \varepsilon_t$, such that $\eta_t^* \sim \mathcal{N}(0, \sigma_\eta^2)$ and $\text{Cov}(\varepsilon_t, \eta_t^*) = 0$. Such parameterization is useful as it ensures Ω is positive-definite for any estimates of σ_ε^2, σ_η^2 and κ_τ.[6]

2.2.2 I(2)-UC models

I(2)-UC models propose an analogous decomposition of y_t as in the I(1) case with one main distinction: the underlying latent level of y_t, τ_t, is augmented by another latent stochastic component, μ_t. Formally, we have:

$$y_t = \tau_t + c_t, \tag{2.2a}$$

$$\tau_t = \mu_t + \tau_{t-1} + \eta_t, \tag{2.2b}$$

$$\mu_t = \mu_{t-1} + \zeta_t, \tag{2.2c}$$

$$\phi(L)c_t = \varepsilon_t, \tag{2.2d}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \\ \zeta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta & \rho_{\varepsilon\zeta}\sigma_\varepsilon\sigma_\zeta \\ \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 & \rho_{\eta\zeta}\sigma_\eta\sigma_\zeta \\ \rho_{\varepsilon\zeta}\sigma_\varepsilon\sigma_\zeta & \rho_{\eta\zeta}\sigma_\eta\sigma_\zeta & \sigma_\zeta^2 \end{bmatrix} \right). \tag{2.2e}$$

Since τ_t is now specified as a random walk with drift process, such models are useful to model variables which grow over time. As a result, instead of inflation, y_t now denotes log price level and τ_t reflects latent movements in *trend price level* rather than trend inflation. Nevertheless, I(2)-UC models can also be perceived as models for inflation. In fact, note that by taking first differences of y_t and defining $\widetilde{c}_t =$

[6] Since we set $\eta_t = \eta_t^* + \kappa_\tau \varepsilon_t$, clearly, with values for σ_ε^2, σ_η^2 and κ_τ, $\rho_{\varepsilon\eta}$ can be recovered using $\rho_{\varepsilon\eta} = \frac{\kappa_\tau \sigma_\varepsilon^2}{\sqrt{\sigma_\varepsilon^2(\kappa_\tau^2 \sigma_\varepsilon^2 + \sigma_\eta^2)}}$.

$(1-L)c_t + \eta_t$, one can reexpress the system above as

$$\Delta y_t = \mu_t + \widetilde{c}_t, \tag{2.3a}$$

$$\mu_t = \mu_{t-1} + \zeta_t, \tag{2.3b}$$

$$\phi(L)\widetilde{c}_t = \phi(L)\eta_t + (1-L)\varepsilon_t, \tag{2.3c}$$

which is essentially the same framework presented earlier for I(1)-UC models except for two facts: (a) the inflation gap, $\widetilde{c}_t$, has now an ARMA (p,q) representation, where $q > 0$; (b) instead of two, I(2)-UC models can accommodate up to three innovations. Consequently, different correlation structures can be explored. Specifically, we look at three approaches to identify correlation amongst ε_t, η_t and ζ_t. As before, we adopt the contrasting SSOE and MSOE schemes whereby all innovations are either perfectly correlated or uncorrelated, but also entertain a new approach which bridges the previous two. We construct UC models which preserve trend inflation (now represented by μ_t) as an orthogonal state but treat MA terms in $\widetilde{c}_t$ as the same stochastic process. In other words, we set $\eta_t = \kappa_\tau \varepsilon_t$. Since this approach represents a midpoint between SSOE and MSOE models, we refer to variants following this identification strategy as reduced source of error (RSOE) models.[7] Below we present the covariance structures respectively associated with the SSOE, RSOE and MSOE schemes:

$$\Omega = \left\{ \begin{bmatrix} \sigma_\varepsilon^2 & \pm\sigma_\varepsilon\sigma_\eta & \pm\sigma_\varepsilon\sigma_\zeta \\ \pm\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 & \pm\sigma_\eta\sigma_\zeta \\ \pm\sigma_\varepsilon\sigma_\zeta & \pm\sigma_\eta\sigma_\zeta & \sigma_\zeta^2 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \pm\sigma_\varepsilon\sigma_\eta & 0 \\ \pm\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\zeta^2 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\zeta^2 \end{bmatrix} \right\}.$$

It should be noted that, if desired, an unrestricted version of Ω for I(2)-UC models could also be estimated. In particular, Oh and Zivot [2006] show that all correlation parameters in (2.2e) can be identified under the likelihood function associated with such models when $\phi(L)c_t$ is specified as an AR(4) process. For parsimony and to avoid potential root cancelation issues we do not pursue such an approach here. Hence, akin to I(1)-UC models we let $p = 0$ and $p = 2$. In the first case we obtain the widely used local linear trend model (see e.g. Harvey and Jaeger [1993], Zarnowitz and Ozyildirim [2006] and Frühwirth-Schnatter and Wagner [2010]). Setting $p = 2$ yields Clark's double-drift UC model (see Clark [1987] and Oh and Zivot [2006]).

[7] For inflation, one possible motivation for such identification scheme is to think of changes in trend inflation as mainly reflecting idiosyncratic systematic changes in the conduct of monetary policy (e.g. Woodford [2007]). Here, broadly captured by ζ_t. On the other hand, ε_t could be perceived as encompassing non-monetary policy factors underlying transitory inflation dynamics. Such transient movements could reflect shifts that affect the (observed or trend) price level but not trend inflation (a plausible scenario when inflation expectations are well-anchored). As an example of such one-off price shifters, one could think of one-off changes in the price level typically observed after changes in taxation (e.g. introduction of value added taxes) and (or) changes in energy and oil prices.

Table 2.1: List of Models

Identifier	Description*
Local Level-SSOE	RW trend inflation and white noise inflation gap; $\rho_{\varepsilon\eta} = \pm 1$
Local Level-MSOE	RW trend inflation and white noise inflation gap; $\rho_{\varepsilon\eta} = 0$
MNZ-SSOE	RW trend inflation and AR(2) inflation gap; $\rho_{\varepsilon\eta} = \pm 1$
MNZ-MSOE(UR)	RW trend inflation and AR(2) inflation gap; $\rho_{\varepsilon\eta} =$ unrestricted
MNZ-MSOE	RW trend inflation and AR(2) inflation gap; $\rho_{\varepsilon\eta} = 0$
Local Linear Trend-SSOE	RW trend inflation and price level; MA(1) inflation gap; $\rho_{\varepsilon\eta} = \rho_{\varepsilon\zeta} = \rho_{\eta\zeta} = \pm 1$
Local Linear Trend-RSOE	RW trend inflation and price level; MA(1) inflation gap; $\rho_{\varepsilon\eta} = \pm 1$ and $\rho_{\varepsilon\zeta} = \rho_{\eta\zeta} = 0$
Local Linear Trend-MSOE	RW trend inflation and price level; MA(1) inflation gap; $\rho_{\varepsilon\eta} = \rho_{\varepsilon\zeta} = \rho_{\eta\zeta} = 0$
CLARK-SSOE	RW trend inflation and price level; ARMA(2,2) inflation gap; $\rho_{\varepsilon\eta} = \rho_{\varepsilon\zeta} = \rho_{\eta\zeta} = \pm 1$
CLARK-RSOE	RW trend inflation and price level; ARMA(2,2) inflation gap; $\rho_{\varepsilon\eta} = \pm 1$ and $\rho_{\varepsilon\zeta} = \rho_{\eta\zeta} = 0$
CLARK-MSOE	RW trend inflation and price level; ARMA(2,2) inflation gap; $\rho_{\varepsilon\eta} = \rho_{\varepsilon\zeta} = \rho_{\eta\zeta} = 0$

*RW stands for random walk.

Finally, as before, combining the covariance structures above with the different orders of p allows us to construct the following six models:

- **Local Linear Trend-SSOE** ($\rho_{\varepsilon\eta} = \pm 1$, $\rho_{\varepsilon\zeta} = \pm 1$, $\rho_{\eta\zeta} = \pm 1$ and $p = 0$): $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \kappa_\mu \varepsilon_t$, $c_t = \varepsilon_t$,

- **Local Linear Trend-RSOE** ($\rho_{\varepsilon\eta} = \pm 1$, $\rho_{\varepsilon\zeta} = \pm 1$, $\rho_{\eta\zeta} = 0$ and $p = 0$): $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $c_t = \varepsilon_t$,

- **Local Linear Trend-MSOE** ($\rho_{\varepsilon\eta} = 0$, $\rho_{\varepsilon\zeta} = 0$, $\rho_{\eta\zeta} = 0$ and $p = 0$): $\tau_t = \mu_t + \tau_t + \eta_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $c_t = \varepsilon_t$,

- **CLARK-SSOE** ($\rho_{\varepsilon\eta} = \pm 1$, $\rho_{\varepsilon\zeta} = \pm 1$, $\rho_{\eta\zeta} = \pm 1$ and $p = 2$): $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \kappa_\mu \varepsilon_t$, $(1 - \phi_1 L - \phi_2 L^2) c_t = \varepsilon_t$,

- **CLARK-RSOE** ($\rho_{\varepsilon\eta} = \pm 1$, $\rho_{\varepsilon\zeta} = \pm 1$, $\rho_{\eta\zeta} = 0$ and $p = 2$): $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $(1 - \phi_1 L - \phi_2 L^2) c_t = \varepsilon_t$,

- **CLARK-MSOE** ($\rho_{\varepsilon\eta} = 0$, $\rho_{\varepsilon\zeta} = 0$, $\rho_{\eta\zeta} = 0$ and $p = 2$): $\tau_t = \mu_t + \tau_t + \eta_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $(1 - \phi_1 L - \phi_2 L^2) c_t = \varepsilon_t$,

where the new loading parameter, κ_μ, in the SSOE variants can be described in a similar fashion as κ_τ in the I(1)-UC case. Table 2.1 summarizes all eleven specifications presented in this section.

2.3 How Changes in the Correlation Structure Can Affect Forecasting

We now discuss how different correlation structures can influence forecasting performance within a Bayesian forecasting framework.[8]

For concreteness, consider now the task of producing k-step ahead forecasts for some variable y_t. Also, let $\mathbf{y} = (y_1, \ldots, y_t)'$ and $\boldsymbol{\theta}^{UC}$ denote an n-dimensional set of parameters associated with any of the UC models described in Section 2.2 such that $\boldsymbol{\theta}^{UC} \in \boldsymbol{\Theta}^{UC} \subseteq \mathbb{R}^n$, where $\boldsymbol{\Theta}^{UC}$ describes the parameter space corresponding to the values of $\boldsymbol{\theta}^{UC}$. A density forecast of y_{t+k} is given as follows:[9]

$$\begin{aligned} f(y_{t+k}|\mathbf{y}) &= \int_{\boldsymbol{\Theta}^{UC}} f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{UC})\ f(\theta^{UC}|\mathbf{y}) d\boldsymbol{\theta}^{UC} \\ &= \int_{\boldsymbol{\Theta}^{UC}} f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{UC})\ \frac{f(\mathbf{y}|\theta^{UC}) f(\boldsymbol{\theta}^{UC})}{f(\mathbf{y})} d\boldsymbol{\theta}^{UC}, \end{aligned} \tag{2.4}$$

where the second term of the integrand in (2.4) follows directly from an application of Bayes' rule. In words, marginalization of $\boldsymbol{\theta}^{UC}$ in $f(y_{t+k}|\mathbf{y})$ implies that Bayesian forecasting accounts for the global properties of the predictive density, $f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{UC})$, and posterior kernel, $f(\mathbf{y}|\boldsymbol{\theta}^{UC}) f(\boldsymbol{\theta}^{UC})$. As such, instead of density forecasts based on a single estimate of $\boldsymbol{\theta}^{UC}$, say, $\hat{\boldsymbol{\theta}}^{UC}$ corresponding to the mode of $f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{UC})$, forecasts from $f(y_{t+k}|\mathbf{y})$ incorporate all possible values of $\boldsymbol{\theta}^{UC}$ within $\boldsymbol{\Theta}^{UC}$. Naturally, since $f(y_{t+k}|\mathbf{y})$ is a function of $\boldsymbol{\Theta}^{UC}$, if the latter is altered then point and density forecast metrics associated with the former are likely to be affected as well. Changes in the correlation between innovations within an UC model can alter $\boldsymbol{\Theta}^{UC}$.

To give an example, consider again the local linear trend model (i.e. when $\phi(L) = 1$) shown in Section 2.2.2. One can readily verify that by taking second differences of the measurement equation in (2.2a) yields:

$$\Delta^2 y_t = \zeta_t + \eta_t - \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}. \tag{2.5}$$

Next, by virtue of Granger's lemma (see Granger and Newbold [1986], p. 28-30), it can be shown that the expression above can be recast as a reduced-form ARIMA

[8] The reader is referred to Geweke and Whiteman [2006] for a detailed discussion on Bayesian forecasting techniques.

[9] For simplicity, we assume $f(.)$ is a continuous probability distribution. Nonetheless, the same ideas underpinning forecasting implications from combining parameter uncertainty and parameter space restrictions are carried over to discrete and mixed discrete-continuous distributions, albeit, involving more cumbersome notation.

(0,2,2) process:

$$\Delta y_t = u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2}, \quad u_t \sim \mathcal{N}(0, \sigma^2), \tag{2.6}$$

where reduced-form parameters, φ_1, φ_2 and σ^2, are nonlinearly related to UC parameters, $\boldsymbol{\theta}^{UC} = \left\{\sigma_\varepsilon^2,\ \sigma_\eta^2,\ \sigma_\zeta^2,\ \rho_{\varepsilon\eta},\ \rho_{\varepsilon\zeta},\ \rho_{\eta\zeta}\right\}$.

It is thus important to recognize that the local linear trend model and its expression in second-difference form in (2.5) represent alternative parameterizations of a reduced form ARIMA (0,2,2) model.[10] As a result, the predictive density in (2.4) can, in principle, also be expressed as:

$$f(y_{t+k}|\mathbf{y}) = \int_{\boldsymbol{\Theta}^{ARIMA}} f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{ARIMA}) \frac{f(\mathbf{y}|\boldsymbol{\theta}^{ARIMA}) f(\boldsymbol{\theta}^{ARIMA})|J|}{f(\mathbf{y})} d\boldsymbol{\theta}^{ARIMA}, \tag{2.7}$$

where $\boldsymbol{\theta}^{ARIMA} = \left\{\varphi_1,\ \varphi_2,\ \sigma^2\right\}$, such that $\boldsymbol{\theta}^{ARIMA} \in \boldsymbol{\Theta}^{ARIMA} \subseteq \mathbb{R}^3$ and $|J|$ denotes the Jacobian of the transformation to parameterize the prior density, $f(\boldsymbol{\theta}^{UC})$, in terms of $\boldsymbol{\theta}^{ARIMA}$. When $|J|$ cannot be computed, an approximation of $f(\boldsymbol{\theta}^{UC})$ in terms of θ^{ARIMA} may be achieved numerically or using techniques such as saddlepoint approximations (see e.g. Goutis and Casella [1999]).

Importantly, the expression in (2.7) suggests that one can study out-of-sample effects of different correlation assumptions on $\rho_{\varepsilon\eta}$, $\rho_{\varepsilon\zeta}$ and $\rho_{\eta\zeta}$ by deriving the restrictions such assumptions imply on the parameter space of a reduced form ARIMA model, $\boldsymbol{\Theta}^{ARIMA}$. To be precise, since $\boldsymbol{\Theta}^{ARIMA}$ is unique (as implied by the Wold decomposition), the predictive likelihood, $f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{ARIMA})$ will take different values for any $\boldsymbol{\theta}^{ARIMA} \in \boldsymbol{\Theta}^{ARIMA}$. Therefore, inasmuch as changes in identifying strategies for $\rho_{\varepsilon\eta}$, $\rho_{\varepsilon\zeta}$ and $\rho_{\eta\zeta}$ alter $\boldsymbol{\Theta}^{ARIMA}$, one can see from (2.7) that $f(y_{t+k}|\mathbf{y})$ will not be invariant to such changes.

To illustrate, Figure 2.1 shows how the SSOE, RSOE and MSOE correlation schemes for the local linear trend generates substantial differences to the support of the predictive likelihood, $f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{ARIMA})$, as measured in terms of the invertibility region of reduced form MA parameters. In particular, looking at the dotted area which describes the support of $\boldsymbol{\theta}^{ARIMA}$ (over the MA parameter space) for the uncorrelated innovations case (i.e. MSOE), one can note that such region is much more restricted relative to the SSOE and RSOE variants. In other words, for the local linear trend model, orthogonality considerably limits the amount of parameter uncertainty

[10] The canonical representation of an ARIMA (0,2,2) model in (2.6) is unique. On the other hand, more than one UC model representation can lead to the same reduced-form representation (see e.g. Cochrane [1988]).

$f(y_{t+k}|\mathbf{y})$ can account for.[11] Obviously, whether or not such constraints are desirable for forecasting is, ultimately, an empirical question. Since RSOE models constitute a new class of models, before turning to estimation, we now state the conditions that ensure that such models yield an invertible ARIMA representation:

Proposition 2.3.1
3.1(a) : *If* $\zeta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\zeta^2)$, *then the Local Linear Trend-RSOE model has an invertible ARIMA(0,2,2) representation.*
3.1(b) : *If the roots of the lag polynomial* $\phi(L)$ *lie outside the unit circle then the CLARK-RSOE model has an invertible ARIMA(2,2,3) representation.*

Proof - See appendix 2.A.1.

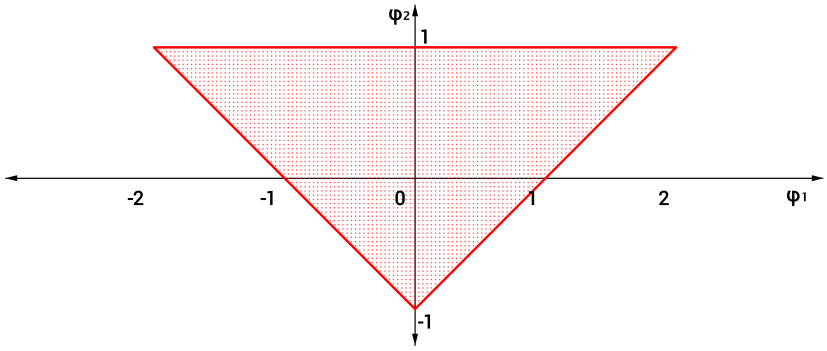
2.4 MCMC Inference of UC Models with Reduced Rank Covariance Matrix

A common approach to estimate UC models (or state space models in general) is to employ MCMC simulation techniques.[12] The usefulness of MCMC sampling in the context of UC models stems from the modular nature of such type of algorithm. In particular, MCMC-based estimation allows one to transform the intractability of direct sampling from a high dimensional joint posterior distribution into a simpler problem of iterative sampling from lower-dimensional conditional posterior distributions. For the SSOE and RSOE variants, however, perfect correlation between some or all innovations reduces the rank of the covariance matrix. As a result, proper MCMC estimation requires addressing matrix singularities which do not occur in the orthogonal and (imperfectly) correlated innovations cases. To the best of our knowledge, MCMC estimation of UC models incorporating the type of rank structures explored in this chapter has not appeared in the literature to date. Therefore, before developing an MCMC algorithm for the models in Section 2.2, it is useful to highlight how such restrictions affect an otherwise standard MCMC sampling scheme.[13]

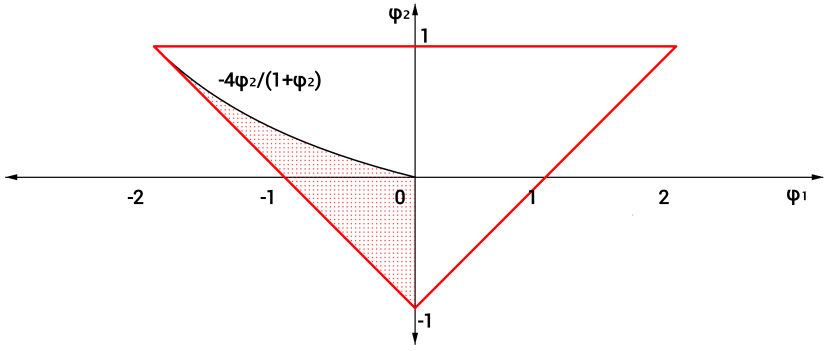
For concreteness, consider again the local linear trend model described in Section 2.2.2.[14] Formal Bayesian estimation of such a model would entail sampling τ_t

[11] Derivation of the admissible regions shown in Figure 2.1 is deferred to Appendix 2.A.3.
[12] The interested reader is referred to e.g. Koop [2003], Gamerman and Lopes [2006] and the references therein for a detailed textbook treatment on MCMC estimation.
[13] Forbes et al. [2000] and Snyder et al. [2001] address Bayesian estimation of state space models with a SSOE representation. Their approach, however, does not encompass MCMC estimation. As is well known, MCMC estimation allows one to work with a wide range of priors. Moreover, once an MCMC algorithm for SSOE and RSOE models is developed, future work could extend such an algorithm to incorporate nonlinearities which accord well with MCMC sampling, such as stochastic volatility.
[14] The discussion above also applies to all the other UC models in Section 2.2

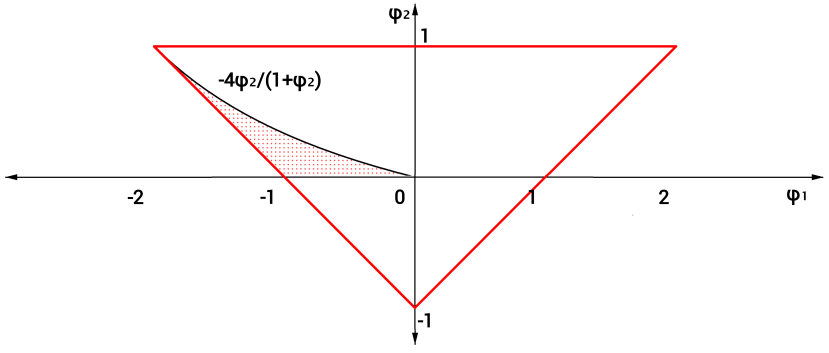
Figure 2.1: Parameter space restrictions over the invertibility region of reduced-form MA parameters following different correlation structures for the local linear trend model. Dotted area denotes the admissible (or non-constrained) region for each correlation structure.



Local Linear Trend-SSOE



Local Linear Trend-RSOE



Local Linear Trend-MSOE

and μ_t through $t = 1, ..., T$ plus parameters (if Ω is unrestricted), σ_ε^2, σ_η^2, σ_ζ^2, $\rho_{\varepsilon\eta}$, $\rho_{\varepsilon\zeta}$ and $\rho_{\eta\zeta}$, from the joint posterior distribution associated with this model. Formally, given the data, $\mathbf{y} = (y_1, ..., y_T)'$, and some initial conditions for the states, τ_0 and μ_0, a common MCMC algorithm for the local linear trend model can be described as a two-step sampling scheme which involves sequentially drawing from the following conditional posteriors:[15]

1. $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})$,
2. $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$,

where $\boldsymbol{\theta} = \{\sigma_\varepsilon^2,\ \sigma_\eta^2,\ \sigma_\zeta^2,\ \rho_{\varepsilon\eta},\ \rho_{\varepsilon\zeta},\ \rho_{\eta\zeta},\ \tau_0,\ \mu_0\}$ and $\mathbf{z} = \{\boldsymbol{\tau},\ \boldsymbol{\mu}\}$, such that $\boldsymbol{\tau} = (\tau_1, ..., \tau_T)'$ and $\boldsymbol{\mu} = (\mu_1, ..., \mu_T)'$.

As is well known, MCMC simulation only requires evaluation of the *kernel* associated with each conditional posteriors above. Thus, to understand how MCMC sampling can be affected by rank reduction of Ω it is useful to address the implications to the kernel of $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})$ and $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$. Proposition 2.4.1 below summarizes such considerations:

Proposition 2.4.1 *Let* $\mathbf{y}$ *and the elements in* $\mathbf{z}$ *denote exchangeable random vectors. If UC models, as the ones considered in Section 2.2, contain one (or more) perfectly correlated state(s) then the kernel of* $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})$ *and* $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$ *exhibits a rank-deficient covariance matrix.*

Proof - See appendix 2.A.2.

While Proposition 2.4.1 might appear intuitive, it has important implications for the design of MCMC samplers of UC models with reduced rank covariance matrix. Specifically, the issues highlighted in Proposition 2.4.1 result from the density degeneracies (i.e. probability distributions with zero variance) which occur when allowing for perfectly correlated states. In such cases, simulation smoothing (i.e. sampling from $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})$) can still be carried out using standard Forward-Filtering-Backward Smoothing (FFBS) algorithms as in Frühwirth-Schnatter [1994], Carter and Kohn [1994], De Jong and Shephard [1995] or Durbin and Koopman [2002]. In particular, as pointed out in, e.g., Harvey and Koopman [2000] and Casals et al. [2015] degeneracies which stem from perfectly correlated states can be handled within the Kalman filter by setting the variance filtering step to zero, which makes the backward

[15] We discuss initialization of state variables in Section 2.5.

smoothing step redundant.[16] In contrast, for parameter sampling (i.e. sampling from $f(\boldsymbol{\theta}|\mathbf{y},\mathbf{z})$), the filtering recursions used for simulation smoothing do not apply and one needs to derive a well-defined density to sample from. This can be achieved by integrating out perfectly correlated states from an MCMC sampler. In the next section we propose a new MCMC sampler which does that by parameterizing UC models in terms of their innovations rather than states.

2.5 Posterior Analysis

In this section we present an efficient posterior simulator for the UC models discussed in Section 2.2. In particular, we develop a disturbance smoothing algorithm which allows us to recover the states, τ_t, μ_t and c_t by simulating the innovations, η_t, ζ_t and ε_t. In doing so, we construct a general estimation framework which readily accommodates the differences in the covariance rank across models. More precisely, UC models are reparameterized in terms of the disturbances driving the state variables τ_t, μ_t and c_t. Within such a framework, MCMC sampling issues discussed in Section 2.4 can be easily tackled. As we show below, once we develop an MCMC algorithm for the UC model with the largest number of disturbances (in our case, the CLARK-MSOE model), estimation of all other UC variants can be treated as special cases nested within a general framework.

Our algorithm differs from other well-known disturbance-smoothing samplers in the literature as in DeJong and Shephard (1995) and Durbin and Koopman (2002). Instead of adopting FFBS recursions we build on recent work in precision-based methods akin to Chan and Jeliazkov [2009] and Chan [2013]. As noted in McCausland et.al (2011), precision-based algorithms are computationally more efficient than their FFBS counterparts. We stress that such computational gains are substantial, especially in recursive forecasting applications (such as ours) exhibiting several models and series which require simulation of posterior distributions, literally, billions of times.

For concreteness, consider now the CLARK-MSOE model discussed in Section 2.2.2. Stacking y_t, τ_t, μ_t and c_t over t for $t = 1, 2..., T$ yields the following matrix

[16] See Harvey [1989] and Durbin and Koopman [2012] for a detailed textbook treatment of the Kalman filter.

representation:

$$\mathbf{y} = \boldsymbol{\tau} + \mathbf{c}, \tag{2.8}$$

$$\mathbf{H}\boldsymbol{\tau} = \boldsymbol{\iota_0}\tau_0 + \boldsymbol{\mu} + \boldsymbol{\eta} \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\eta), \tag{2.9}$$

$$\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} \qquad \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\zeta) \tag{2.10}$$

$$\mathbf{H}_\phi\mathbf{c} = \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\varepsilon), \tag{2.11}$$

where $\boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_T$ for $i = \varepsilon,\ \eta$ and ζ; $\boldsymbol{\tau} = (\tau_1, ...\tau_T)'$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_T)'$, $\mathbf{c} = (c_1, ..., c_T)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$, $\boldsymbol{\eta} = (\eta_1, ...\eta_T)'$, $\boldsymbol{\zeta} = (\zeta_1, ...\zeta_T)'$, $\boldsymbol{\iota_0} = (1, 0, ..., 0)'$ and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, \ \mathbf{H}_\phi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\phi_1 & 1 & 0 & 0 & 0 \\ -\phi_2 & -\phi_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & -\phi_2 & -\phi_1 & 1 \end{bmatrix}.$$

Two comments are in order here. First, note that both $\mathbf{H}$ and $\mathbf{H}_\phi$ are banded $T \times T$ matrices. More specifically, they are lower triangular Toeplitz matrices. In what follows, we explore the sparse structure and the commutative property associated with such matrices (see e.g. Pollock et al. [1999], p. 644) to develop a disturbance-based parametrization of the system above which enables fast posterior simulation. Second, akin to Snyder et al. [2001] we initialize τ_t and μ_t using Winter's approach (see Winters [1960]) to construct the initial conditions, τ_0 and μ_0, based on the first five years of data. Also, for simplicity, pre-sample values for the stationary state, c_t, are set to zero. Therefore, initial conditions are treated as predetermined terms and do not enter the MCMC sampling algorithm discussed below.[17]

Next, we derive a disturbance-based parameterization of the model described in

[17] If desired, one could treat τ_0, μ_0, c_0 and c_{-1} as parameters, hence augmenting our MCMC algorithm to draw from the conditional posteriors of such parameters. Winters' exponential smoothing method, on the other hand, is easy to implement and reduces the number of parameters one needs to sample. Also, the choice of initialization for the states is unlikely to significantly affect forecast performance. For additional approaches to initialize state space models which share similarities with the framework presented here, the reader is referred to De Jong and Chu-Chun-Lin [1994] and Casals and Sotoca [2001].

(2.8)-(2.11). To do so, notice that by pre-multiplying both sides of (2.8) by **HH** gives:

$$\mathbf{HHy} = \mathbf{HH}\boldsymbol{\tau} + \mathbf{HHc} \tag{2.12a}$$

$$\mathbf{HHy} = \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \mathbf{H}\boldsymbol{\mu} + \mathbf{H}\boldsymbol{\eta} + \mathbf{HHc} \tag{2.12b}$$

$$(\mathbf{HH})^{-1}\mathbf{H}_\phi\mathbf{HHy} = (\mathbf{HH})^{-1}\mathbf{H}_\phi\left(\mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \boldsymbol{\zeta} + \mathbf{H}\boldsymbol{\eta} + \mathbf{HHc}\right) \tag{2.12c}$$

$$\widetilde{\mathbf{y}} = \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \tag{2.12d}$$

where $\widetilde{\mathbf{y}} = \mathbf{H}_\phi\mathbf{y}$, $\mathbf{z_0} = (\tau_0 \;\; \mu_0)'$, $\mathbf{X_0}$ is a $T \times 2$ matrix, $\mathbf{X_0} = ((\mathbf{HH})^{-1}\mathbf{H}_\phi\mathbf{H}\boldsymbol{\iota}_0 \;\; (\mathbf{HH})^{-1}\mathbf{H}_\phi\boldsymbol{\iota}_0)$, and $\mathbf{X_1}$ and $\mathbf{X_2}$ represent $T \times T$ matrices defined as $\mathbf{X_1} = \mathbf{X_2} = \mathbf{H}_\phi$. The disturbance vectors $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$, by a simple change of variable, are now denoted as $\widetilde{\boldsymbol{\eta}} = \mathbf{H}^{-1}\boldsymbol{\eta}$ and $\widetilde{\boldsymbol{\zeta}} = (\mathbf{HH})^{-1}\boldsymbol{\zeta}$. Once a draw for $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\zeta}}$ is obtained, the original disturbance vectors can be readily recovered using $\boldsymbol{\eta} = \mathbf{H}\widetilde{\boldsymbol{\eta}}$ and $\boldsymbol{\zeta} = \mathbf{HH}\widetilde{\boldsymbol{\zeta}}$. Note also that the specification in (2.12d) is possible since $(\mathbf{HH})^{-1}$ and $\mathbf{H}_\phi$ are lower triangular Toeplitz matrices. Therefore, using $(\mathbf{HH})^{-1}\mathbf{H}_\phi = \mathbf{H}_\phi(\mathbf{HH})^{-1}$ in (2.12c) it is easy to verify that (2.12d) ensues.[18]

Now recall that once we present the measurement equation as in (2.12d), the disturbances, $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\zeta}}$, can be interpreted as our new state vectors. As a result, the state-space representation in (2.8)-(2.11) can be recast as:

$$\widetilde{\mathbf{y}} = \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\varepsilon), \tag{2.13}$$

$$\widetilde{\boldsymbol{\eta}} \sim \mathcal{N}(\mathbf{0},\ \mathbf{D}_{\widetilde{\eta}}), \tag{2.14}$$

$$\widetilde{\boldsymbol{\zeta}} \sim \mathcal{N}(\mathbf{0},\ \mathbf{D}_{\widetilde{\zeta}}), \tag{2.15}$$

where $\mathbf{D}_{\widetilde{\eta}} = \mathbf{H}^{-1}\boldsymbol{\Sigma}_\eta\mathbf{H}'^{-1}$ and $\mathbf{D}_{\widetilde{\zeta}} = (\mathbf{HH})^{-1}\boldsymbol{\Sigma}_\zeta(\mathbf{HH})'^{-1}$.

The usefulness of parameterizing UC models in terms of innovations rather than states (i.e. τ_t, μ_t and c_t) becomes more evident in the case of UC models which contain a covariance matrix with reduced rank. In particular, since rank reduction stems from reducing the number of innovations, the representation in (2.13)-(2.15) provides an intuitive way to think of SSOE and RSOE schemes as nested cases within an 'innovations-richer' framework. For example, in the SSOE case one could first set $\boldsymbol{\eta} = \kappa_\tau\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta} = \kappa_\mu\boldsymbol{\varepsilon}$ in (2.12b) and apply straightforward algebraic manipulations to derive a measurement equation as in (2.12d), except that now $\mathbf{X_1} = \mathbf{X_2} = \mathbf{0}_{(T\times T)}$ since SSOE models do not require sampling $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\zeta}}$. Naturally, $\widetilde{\mathbf{y}}$ and $\mathbf{X_0}$ would also change accordingly.

In fact, all UC variants entertained in this chapter can be accommodated into

[18] We also use the fact that the product between and the inverse of two lower triangular Toeplitz matrices yield another lower triangular Toeplitz matrix. In other words, if **H** is a lower triangular Toeplitz matrix so are **HH** and $(\mathbf{HH})^{-1}$.

the framework above for appropriately defined $\widetilde{\mathbf{y}}$, $\widetilde{\boldsymbol{\eta}}$, $\widetilde{\boldsymbol{\zeta}}$, $\mathbf{X_0}$, $\mathbf{X_1}$, $\mathbf{X_2}$ and $\mathbf{z_0}$. Doing so reduces the coding burden typically associated with adapting an MCMC algorithm to various problems.[19] To avoid cluttering the discussion here with algebraic details we defer the derivation and presentation of the exact structures of such matrices and vectors for each model to a technical appendix.

Obtaining posterior draws for the representation in (2.13)-(2.15) can, thus, be summarized as a three-step algorithm which requires sequentially sampling from:[20]

1. $f(\widetilde{\boldsymbol{\eta}}|\mathbf{y}, \mathbf{z}_{-\widetilde{\eta}}, \boldsymbol{\theta})$,
2. $f(\widetilde{\boldsymbol{\zeta}}|\mathbf{y}, \mathbf{z}_{-\widetilde{\zeta}}, \boldsymbol{\theta})$,
3. $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$,

where we adopt the notation $\mathbf{z}_{-\mathbf{j}}$ to describe elements in $\mathbf{z}$ other than $\mathbf{j}$.

Steps 1 and 2 represent the disturbance smoothing block, step 3 denotes parameter (block) sampling. In practice, reducing the number of innovations entails removing steps 1 and 2 from the MCMC algorithm accordingly. In other words, depending on the correlation structure one can have $\mathbf{z} = \{\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\zeta}}\}$ (i.e. MSOE case), $\mathbf{z} = \{\widetilde{\boldsymbol{\zeta}}\}$ (i.e. RSOE case) or $\mathbf{z} = \emptyset$ (i.e. SSOE case). Similarly, parameters in $\boldsymbol{\theta}$ are also model contingent. For the CLARK-MSOE model we have: $\boldsymbol{\theta} = \{\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \phi_1, \phi_2\}$. Nonetheless, despite the different configurations of $\boldsymbol{\theta}$, parameter draws are obtained using the same strategy across models, namely a Metropolis-within-Gibbs algorithm. Parameter sampling is discussed in Section 2.5.2. We turn next to the discussion of disturbance smoothing.

2.5.1 Disturbance Smoothing

This section introduces a direct and efficient way to sample $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ required for the MSOE and RSOE schemes. We begin by sampling $\boldsymbol{\eta}$. To do so, note first that since $\boldsymbol{\varepsilon}$ and $\widetilde{\boldsymbol{\eta}}$ are normally distributed random vectors, the conditional likelihood, $f(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta})$, and prior $f(\widetilde{\boldsymbol{\eta}}|\boldsymbol{\theta})$ can be expressed as:

$$f(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_\varepsilon|^{-\frac{1}{2}} \exp\left(-\frac{(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}})' \boldsymbol{\Sigma}_\varepsilon^{-1} (\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}})}{2}\right), \tag{2.16}$$

[19] More specifically, adopting general notation as in (2.13)-(2.15) enables one to describe posterior moments for all models in terms of common matrix structures.

[20] For notational convenience we exclude $\mathbf{z_0}$ as a conditioning factor in conditional posteriors since initial conditions are predetermined using Winter's approach.

and

$$f(\widetilde{\boldsymbol{\eta}}|\boldsymbol{\theta}) \propto |\mathbf{D}_{\widetilde{\eta}}|^{-\frac{1}{2}} \exp\left(-\frac{\widetilde{\boldsymbol{\eta}}'\mathbf{D}_{\widetilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}}}{2}\right), \tag{2.17}$$

where $\widetilde{\mathbf{y}}_{\tilde{\eta}} = \widetilde{\mathbf{y}} - \mathbf{X_0 z_0} - \mathbf{X_1}\widetilde{\boldsymbol{\zeta}}$.

Next, using Bayes' rule to combine (2.16) with (2.17) and applying standard regression results (e.g. Koop et al. [2007]) yields:

$$\begin{aligned} f(\widetilde{\boldsymbol{\eta}}|\mathbf{y}, \mathbf{z}_{-\tilde{\mathbf{j}}}, \boldsymbol{\theta}) &\propto \exp\left(-\frac{(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}})'\boldsymbol{\Sigma}_\varepsilon^{-1}(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}}) + \widetilde{\boldsymbol{\eta}}'\mathbf{D}_{\widetilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}}}{2}\right) \\ &\propto \exp\left(-\frac{\widetilde{\boldsymbol{\eta}}'\left(\mathbf{X_2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\widetilde{\eta}}^{-1}\right)\widetilde{\boldsymbol{\eta}} - 2\widetilde{\mathbf{y}}_{\tilde{\eta}}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2}\widetilde{\boldsymbol{\eta}}}{2}\right) \\ &= \exp\left(-\frac{\widetilde{\boldsymbol{\eta}}'\overline{\mathbf{D}}_{\widetilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}} - 2\bar{\mathbf{d}}_{\tilde{\eta}}'\overline{\mathbf{D}}_{\widetilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}}}{2}\right). \end{aligned}$$

The expression above reveals a Gaussian kernel for $(\widetilde{\boldsymbol{\eta}}|\mathbf{y}, \mathbf{z}_{-\tilde{\eta}}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\tilde{\eta}}, \overline{\mathbf{D}}_{\tilde{\eta}}\right)$, where $\overline{\mathbf{D}}_{\tilde{\eta}} = \left(\mathbf{X_2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\widetilde{\eta}}^{-1}\right)^{-1}$ and $\bar{\mathbf{d}}_{\tilde{\eta}} = \overline{\mathbf{D}}_{\tilde{\eta}}\mathbf{X_2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\widetilde{\mathbf{y}}_{\tilde{\eta}}$. Now, remember that we defined $\mathbf{X_2} = \mathbf{H}_\phi$ and $\mathbf{D}_{\tilde{\mathbf{j}}} = \mathbf{H}^{-1}\boldsymbol{\Sigma}_\eta\mathbf{H}'^{-1}$. Using these two results, it is easy to verify that the precision matrix, $\overline{\mathbf{D}}_{\tilde{\eta}}^{-1} = \left(\mathbf{H}_\phi'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{H}_\phi + \mathbf{H}'\boldsymbol{\Sigma}_\eta^{-1}\mathbf{H}\right)$ is a sparse matrix with a penta-diagonal structure. To be precise, this means $\overline{\mathbf{D}}_{\tilde{\eta}}^{-1}$ contains $5T - 6$ non-zero entries, which is substantially less than T^2 non-zero entries as in the case of full $T \times T$ matrix. As a result, we can implement the precision sampler of Chan and Jeliazkov [2009] which exploits the banded structure of $\overline{\mathbf{D}}_{\tilde{\eta}}^{-1}$ to expedite computation. In particular, the authors show how $\overline{\mathbf{D}}_{\tilde{\eta}} = \left(\mathbf{X_2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\widetilde{\eta}}^{-1}\right)^{-1}$ can be computed using three steps of $\mathcal{O}(T)$ operations instead of $\mathcal{O}(T^3)$ operations, which is what is required if computing $\left(\mathbf{X_2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\widetilde{\eta}}^{-1}\right)^{-1}$ via brute-force inversion (see e.g. Golub and Van Loan [1983] p.156).

To illustrate how we adapt their algorithm for disturbance smoothing, we introduce the following notation: given a lower (upper) triangular $T \times T$ non-singular matrix $\mathbf{C}$ and a $T \times 1$ vector $\mathbf{b}$, let $\mathbf{C} \setminus \mathbf{b}$ denote the unique solution to the triangular system $\mathbf{Cx} = \mathbf{b}$ obtained by forward (backward) substitution, i.e., $\mathbf{C} \setminus \mathbf{b} = \mathbf{C}^{-1}\mathbf{b}$. That said, sampling $(\widetilde{\boldsymbol{\eta}}|\mathbf{y}, \mathbf{z}_{-\tilde{\eta}}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\tilde{\eta}}, \overline{\mathbf{D}}_{\tilde{\eta}}\right)$ is then conducted following four

$\mathcal{O}(T)$ operations:

$$
\begin{aligned}
&(1) \quad Chol(\overline{\mathbf{D}}_{\widetilde{\eta}}^{-1}) = \mathbf{C}\mathbf{C}', \\
&(2) \quad \mathbf{x} = \mathbf{C} \setminus \left(\mathbf{X}_2' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \widetilde{\mathbf{y}}_{\widetilde{\eta}} \right), \\
&(3) \quad \overline{\mathbf{d}}_{\widetilde{\eta}} = \mathbf{C}' \setminus \mathbf{x}, \\
&(4) \quad \widetilde{\boldsymbol{\eta}} = \overline{\mathbf{d}}_{\widetilde{\eta}} + \mathbf{C}' \setminus \mathbf{u} \qquad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).
\end{aligned}
$$

The first step describes the Cholesky decomposition of $\overline{\mathbf{D}}_{\widetilde{\eta}}^{-1}$, such that $\overline{\mathbf{D}}_{\widetilde{\eta}}^{-1} = \mathbf{C}\mathbf{C}'$. Since $\overline{\mathbf{D}}_{\widetilde{\eta}}^{-1}$ is a banded matrix, a Cholesky factorization only involves $\mathcal{O}(T)$ operations (see Golub and Van Loan [1983] p.156). Step 2 requires solving a triangular system by forward substitution (given that $\mathbf{C}$ is a lower triangular matrix) which entails $\mathcal{O}(T)$ operations as well. Step 3 is equivalent to Step 2, except that the solution of the triangular system, $\mathbf{C}' \setminus \mathbf{x}$, is now obtained by backward substitution. It is then straightforward to see that Steps 2 and 3 combined, by definition, yield

$$
\overline{\mathbf{d}}_{\widetilde{\eta}} = \mathbf{C}'^{-1} \left(\mathbf{C}^{-1} \left(\mathbf{X}_2' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \widetilde{\mathbf{y}}_{\widetilde{\eta}} \right) \right) = \left(\mathbf{C}\mathbf{C}' \right)^{-1} \left(\mathbf{X}_2' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \widetilde{\mathbf{y}}_{\widetilde{\eta}} \right) = \overline{\mathbf{D}}_{\widetilde{\eta}}^{-1} \left(\mathbf{X}_2' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \widetilde{\mathbf{y}}_{\widetilde{\eta}} \right).
$$

Finally, Step 4 describes an affine transformation of standard normal random vector $\mathbf{u}$. Hence, by sampling T independent standard normal draws $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, one can readily verify that the last step in the algorithm above returns a $T \times 1$ random vector $\widetilde{\boldsymbol{\eta}} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\widetilde{\eta}}, \overline{\mathbf{D}}_{\widetilde{\eta}} \right)$. As mentioned earlier, once we obtained $\widetilde{\boldsymbol{\eta}}$, one can check how the latter is parameterized to recover $\boldsymbol{\eta}$. In the case of the CLARK-MSOE model we have $\widetilde{\boldsymbol{\eta}} = \mathbf{H}^{-1}\boldsymbol{\eta}$, hence $\boldsymbol{\eta} = \mathbf{H}\widetilde{\boldsymbol{\eta}}$.

Posterior simulation of $f(\widetilde{\boldsymbol{\zeta}}|\mathbf{y}, \mathbf{z}_{\_\widetilde{\zeta}}, \boldsymbol{\theta})$ can be carried out just as described for $f(\widetilde{\boldsymbol{\eta}}|\mathbf{y}, \mathbf{z}_{\_\widetilde{\eta}}, \boldsymbol{\theta})$, except that now one needs to combine the likelihood in (2.16) with the following prior density:

$$
f(\widetilde{\boldsymbol{\zeta}}|\boldsymbol{\theta}) \propto |\mathbf{D}_{\widetilde{\zeta}}|^{-\frac{1}{2}} \exp\left(-\frac{\widetilde{\boldsymbol{\zeta}}' \mathbf{D}_{\widetilde{\zeta}}^{-1} \widetilde{\boldsymbol{\zeta}}}{2} \right).
$$

For the sake of brevity, we skip redundant algebraic manipulations and directly present posterior moments associated with $f(\widetilde{\boldsymbol{\zeta}}|\mathbf{y}, \mathbf{z}_{\_\widetilde{\zeta}}, \boldsymbol{\theta})$. Formally, we have $(\widetilde{\boldsymbol{\zeta}}|\mathbf{y}, \mathbf{z}_{\_\widetilde{\zeta}}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\widetilde{\zeta}}, \overline{\mathbf{D}}_{\widetilde{\zeta}} \right)$, where $\overline{\mathbf{D}}_{\widetilde{\zeta}} = \left(\mathbf{X}_1' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \mathbf{X}_1 + \mathbf{D}_{\widetilde{\zeta}}^{-1} \right)^{-1}$ and $\overline{\mathbf{d}}_{\widetilde{\zeta}} = \overline{\mathbf{D}}_{\widetilde{\zeta}} \mathbf{X}_2' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \widetilde{\mathbf{y}}_{\widetilde{\zeta}}$, where $\widetilde{\mathbf{y}}_{\widetilde{\zeta}} = \widetilde{\mathbf{y}} - \mathbf{X_0}\mathbf{z_0} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}}$. Note that $\mathbf{X_1} = \mathbf{H}_{\phi}$, $\mathbf{X}_1' \boldsymbol{\Sigma}_{\varepsilon}^{-1} \mathbf{X_1}$ and $\mathbf{D}_{\widetilde{\zeta}}^{-1} = (\mathbf{H}\mathbf{H})' \boldsymbol{\Sigma}_{\zeta}^{-1} (\mathbf{H}\mathbf{H})$ are all sparse matrices. Therefore, draws of $\widetilde{\boldsymbol{\zeta}} = (\mathbf{H}\mathbf{H})^{-1}\boldsymbol{\zeta}$ can also be quickly obtained following the four $\mathcal{O}(T)$ steps described above. Finally, by setting $\boldsymbol{\zeta} = \mathbf{H}\mathbf{H}\widetilde{\boldsymbol{\zeta}}$ allows one to recover $\boldsymbol{\zeta}$.

2.5.2 Parameter Sampling

Just like state innovations, parameters are also specification contingent. For example, recall that the loading parameters, κ_τ and κ_μ only appear in UC models containing two or more correlated states. Similarly, ϕ_1 and ϕ_2 only need to be sampled in UC models where c_t has an autoregressive representation. Despite such specificities, parameter sampling can be described within a general framework as well.

To be clear, let $\boldsymbol{\theta} = \{\sigma^2_\varepsilon,\ \sigma^2_\eta,\ \sigma^2_\zeta,\ \kappa_\tau,\ \kappa_\mu,\ \phi_1,\ \phi_2\}$ denote the set containing all possible parameters for any of the UC models discussed in Section 2.2. Moreover, let $\boldsymbol{\theta}_{-\sigma^2}$ denote any specification consistent subset of parameters in $\boldsymbol{\theta}$ such that variance parameters, σ^2_ε, σ^2_η and σ^2_ζ, are excluded. Similarly, let $\boldsymbol{\sigma^2}$ denote a subset of $\boldsymbol{\theta}$ containing specification consistent variance parameters. To illustrate, in the case of the CLARK-MSOE model we have $\boldsymbol{\theta}_{-\sigma^2} = \{\phi_1,\ \phi_2\}$ and $\boldsymbol{\sigma^2} = \{\sigma^2_\varepsilon,\ \sigma^2_\eta,\ \sigma^2_\zeta\}$. An MCMC sampling scheme for the elements in $\boldsymbol{\theta}$ can, thus, be recast as a two-step algorithm:

3.1. $f(\boldsymbol{\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}_{-\sigma^2})$,
3.2. $f(\boldsymbol{\theta}_{-\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$.

To sample from the distributions above we consider the following independent priors:[21],

$$\sigma^2_i \sim \mathcal{IG}(\underline{\nu}_i,\ \underline{\mathcal{S}}_i) \text{ for } i = \varepsilon,\ \eta \text{ and } \zeta; \quad \kappa_i \sim \mathcal{N}(0, \underline{\sigma}^2_i)\mathbb{I}_{(\boldsymbol{\psi}\in A_{\boldsymbol{\psi}})} \text{ for } i = \tau \text{ and } \mu;$$

$$\phi_i \sim \mathcal{N}(0, \underline{\sigma}^2_{\phi_i})\mathbb{I}_{(\boldsymbol{\phi}\in A_{\boldsymbol{\phi}})} \text{ for } i = 1 \text{ and } 2,$$

where $\mathcal{IG}$ denotes an inverse-gamma density and $\mathbb{I}_{(\boldsymbol{\psi}\in A_{\boldsymbol{\psi}})}$ and $\mathbb{I}_{(\boldsymbol{\phi}\in A_{\boldsymbol{\phi}})}$, respectively, represent indicator functions which ensure draws of κ_i and ϕ_i are compatible with an invertible and stationary reduced form ARIMA representation of the UC models in Section 2.2. The exact invertibility ($A_{\boldsymbol{\psi}}$) and stationary $A_{\boldsymbol{\phi}}$ conditions required for the UC models adopted in this chapter, are discussed on a case-by-case basis in Appendix 2.A.4.

In practice, posterior draws from $f(\boldsymbol{\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}_{-\sigma^2})$ can be obtained by sampling each variance parameter separately from an inverse-gamma density, i.e., the variance parameters in $\boldsymbol{\sigma^2}$ are a posteriori independent. In fact, using standard methods (see e.g. Koop [2003]) one can verify that

$$\sigma^2_i|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}_{-\sigma^2_i} \sim \mathcal{IG}\left(\underline{\nu}_i + \frac{T}{2},\ \bar{\mathcal{S}}_i\right) \quad \text{for } i = \varepsilon,\ \eta \text{ and } \zeta,$$

[21] Prior hyperparameters are presented in Section 2.6.1

where $\bar{\mathcal{S}}_i = \underline{\mathcal{S}}_i + \frac{\sum_{t=1}^{T} i_t^2}{2}$ for $i = \varepsilon,\ \eta$ and ζ.

Next, draws of the loading and autoregressive parameters require simulating $f(\boldsymbol{\theta}_{-\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$. The latter, however, is not of a known form which one can readily sample from. To circumvent this issue, we introduce a Metropolis-Hastings step to our algorithm. To do so, note first that combining the likelihood function in (2.16) with a joint prior density $f(\boldsymbol{\theta}_{-\sigma^2})$ yields:

$$\log f(\boldsymbol{\theta}_{-\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2}) \propto \log\left(-\frac{\widetilde{\mathbf{y}}_*' \boldsymbol{\Sigma}_\varepsilon^{-1} \widetilde{\mathbf{y}}_*}{2}\right) + \log f(\boldsymbol{\theta}_{-\sigma^2}),$$

where $\widetilde{\mathbf{y}}_* = \widetilde{\mathbf{y}} - \mathbf{X_0 z_0} - \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}}$. Notably, depending on the UC model specification, parameters in $\boldsymbol{\theta}_{-\sigma^2}$ show up in different components of the measurement equation in (2.13). In the CLARK-MSOE model, for example, recall from our previous discussion that $\boldsymbol{\theta}_{-\sigma^2} = \{\phi_1,\ \phi_2\}$, $\mathbf{X_0} = ((\mathbf{HH})^{-1}\mathbf{H}_\phi \mathbf{H}\boldsymbol{\iota}_0 \quad (\mathbf{HH})^{-1}\mathbf{H}_\phi \boldsymbol{\iota}_0)$, and $\mathbf{X_1} = \mathbf{X_2} = \mathbf{H}_\phi$.[22] Nonetheless, regardless of the specification all such matrices are banded. Hence, $\log f(\boldsymbol{\theta}_{-\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$ can be quickly evaluated using sparse routines and the generalized inverse operator ($\backslash$) discussed in Section 2.5.1, both implemented in most statistical packages. Then, draws from $f(\boldsymbol{\theta}_{-\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$ are obtained using an independence-chain Metropolis-Hastings step (see, e.g., Tierney [1994]) with proposal density given by $\mathcal{N}\left(\hat{\boldsymbol{\theta}}_{-\sigma^2},\ \mathbf{G}^{-1}\right)$ where a Newton-Raphson method is adopted to numerically compute the mode and the negative Hessian evaluated at the mode of $f(\boldsymbol{\theta}_{-\sigma^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$, denoted as $\hat{\boldsymbol{\theta}}_{-\sigma^2}$ and $\mathbf{G}^{-1}$, respectively.

2.6 Evaluation

In this section we empirically evaluate the effects to forecasting performance that stem from allowing for different state correlation structures. Even though our focus is on forecasting, we also present results for trend inflation measures that arise from different correlation structures as well as computational efficiency results for the MCMC algorithm developed in Section 2.5.

2.6.1 Data and Priors

Our data set consists of quarterly series for CPI and the implicit price deflators for real GDP and personal consumption expenditure (PCE). Quarterly series were constructed by averaging monthly data over the 3 months within the quarter. Given a

[22] Again, the interested reader is referred to technical appendix 2.A.5, for a detailed description of the exact structures underlying the matrices in the disturbance-based parameterization of all the other UC models employed in this chapter.

Table 2.2: Priors

Model	σ_ε^2	σ_η^2	σ_ξ^2	κ_τ	κ_μ	ϕ_1	ϕ_2
Local Level-SSOE	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)$	–	–	–
Local Level-MSOE	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9)$	–	–	–	–	–
MNZ-SSOE	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$
MNZ-MSOE(UR)	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9)$	–	$\mathcal{N}(0,0.49)$	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$
MNZ-MSOE	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9)$	–	–	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$
Local Linear Trend-SSOE	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\phi)}$	–	–
Local Linear Trend-RSOE	$\mathcal{IG}(10,9)$	–	$\mathcal{IG}(10,9)$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	–	–	–
Local Linear Trend-MSOE	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9^{-10^6})$	$\mathcal{IG}(10,9)$	–	–	–	–
CLARK-SSOE	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\psi\in A_\phi)}$
CLARK-RSOE	$\mathcal{IG}(10,9)$	–	$\mathcal{IG}(10,9)$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$
CLARK-MSOE	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9^{-10^6})$	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$

quarterly price level figure, x_t, we use $y_t = 400\log(x_t/x_{t-1})$ to capture price changes at a quarterly annualized basis. Again, remember that for I(1)-UC models y_t denotes the first difference in logs of a price level measure while for I(2)-UC models y_t denotes the actual log-price level measure at a quarterly annualized basis, i.e, $y_t = 400\log(x_t)$. Figure 2.2 shows the data used in our empirical exercise.

Priors are selected to balance three criteria: facilitate comparison across models, being relatively uninformative and follow recommendations from previous studies. For example, in keeping with previous forecasting studies (e.g., Stock and Watson [2007], Chan [2013] and Clark and Doh [2014]) we follow the practice of using inverse-gamma priors for variance parameters with hyperparameters calibrated to reflect reasonably uninformative priors. An exception to that is σ_η^2 for I(2)-UC models. In particular, we follow the recommendation in Zarnowitz and Ozyildirim [2006] who suggest filtering I(2) processes with an UC representation that assigns quite small conditional variance of the trend level state (i.e., τ_t).[23] In addition, for the AR coefficients, we follow Garnier et al. [2015] who parameterize the joint prior density of ϕ_1 and ϕ_2 tightly around zero to forecast inflation. Table 2.2 summarizes prior hyperparameters for each model.

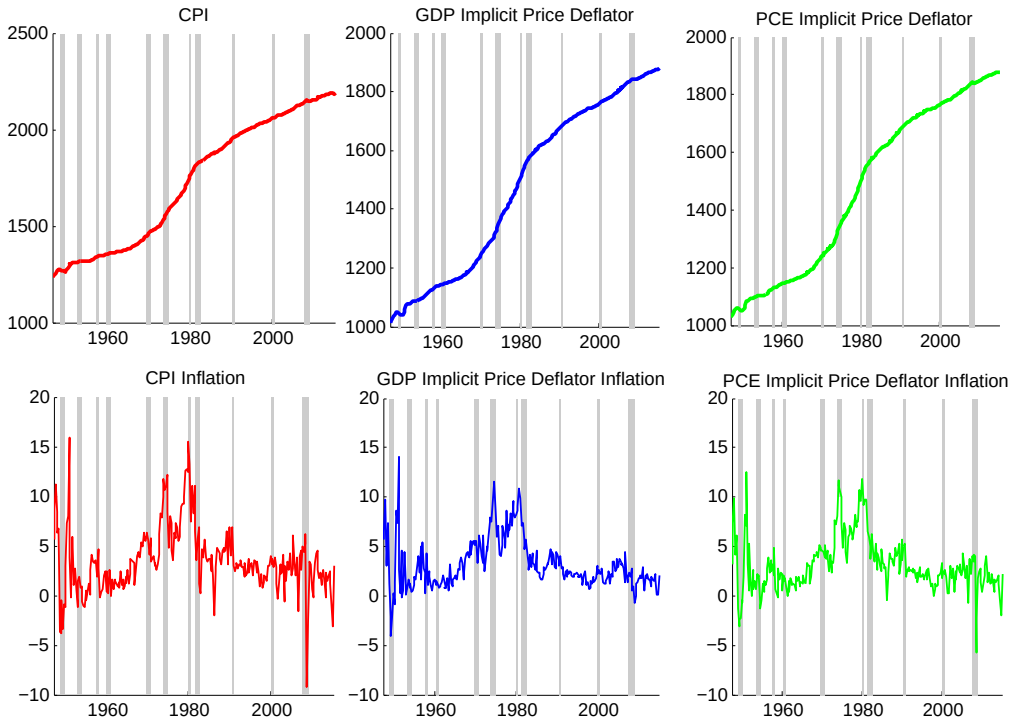
2.6.2 The Forecasting Algorithm

We now use all models listed in Table 2.1 to carry out a recursive forecasting exercise for the series in Figure 2.2.[24] Forecasts are generated for the periods from 1971Q1 through 2015Q2 and assessed on their k-step-ahead point and density prediction performance, for $k = 1, 2, 4, 8, 12, 16$. To measure point forecast accuracy we compute the root mean square forecast error (RMSFE) associated with each model. Density forecasts are evaluated in terms of average predictive log-scores.

[23] In fact, such an assumption leads to an UC model which approximates a parametric representation of the widely used (non-parametric) Hodrick-Prescott filter (see Harvey and Jaeger [1993] for details).

[24] See Marcellino et al. [2006] for details on the benefits of using iterated rather than direct step-ahead forecasting methods.

Figure 2.2: US Quarterly Measures of Annualized Price Level and Inflation from 1947Q1 to 2015Q2



Note: Shaded regions indicate recessions as recorded by the NBER

To gauge the statistical significance of the differences in forecasting performance and in keeping with recent studies (see, e.g., Bauwens et al. [2014], Clark and Doh [2014], Clark and Ravazzolo [2014] and Garnier et al. [2015]) we report results for the Diebold and Mariano [1995] t-test based on a quadratic loss function which, under the null hypothesis, postulates equivalent forecasting accuracy between competing models. To control for serial correlation in forecast errors, as in Clark and Doh [2014], standard errors of t-statistics are computed using a heteroskedasticity and autocorrelation-consistent (HAC) with pre-whitened quadratic spectral estimator. Such test is applied for both RMSFE and predictive log-score results.[25]

To generate forecasts we conduct a predictive simulation exercise along the lines of Cogley et al. [2005] adapted to the disturbance-based parametrization discussed in Section 2.5. To illustrate, let y_t denote the log of a price level measure, such that Δy_{t+k} denotes the k-step ahead *inflation* forecast (i.e. $y_{t+k} - y_{t+k-1}$). Also, let $\mathbf{y_{1:t}}$ and $\mathbf{z_{1:t}} = \{\boldsymbol{\eta}_{1:t},\ \boldsymbol{\zeta}_{1:t}\}$ denote vectors containing data and state innovations up to time t, respectively. Therefore, akin to the discussion in Section 2.3 and using standard results for conditional probability, a k-step ahead predictive density can be expressed as:

$$f(\Delta y_{t+k}|\mathbf{y_{1:t}}) = \int_{\mathcal{F}} \prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta}) f(\mathbf{z_{t+1:t+k}}|\mathbf{y_{1:t}}, \mathbf{z_{1:t}}, \boldsymbol{\theta}) f(\mathbf{z_{1:t}}, \boldsymbol{\theta}|\mathbf{y_{1:t}}) d\mathcal{F},$$

where

$$\mathcal{F} = \begin{cases} \{\mathbf{z_{1:t+k}}, \boldsymbol{\theta}\} & \text{if } k = 1, \\ \{\Delta\mathbf{y_{t+1:t+k-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta}\} & \text{if } k > 1. \end{cases}$$

In practice, however, it is not possible to analytically evaluate the high-dimensional integral above. A common approach to circumvent this issue is to apply Monte Carlo integration techniques to approximate $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$ numerically. Specifically, a central limit theorem can be evoked (see, e.g., Geweke [1992]) to generate the following

[25] See, e.g., Clark and Doh [2014], Clark and Ravazzolo [2014] and Garnier et al. [2015] for recent applications of the Diebold and Mariano [1995] test to assess differences in density forecasting performance.

Rao-Blackwellized estimator of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$:

$$\widehat{f}(\Delta y_{t+k}|\mathbf{y_{1:t}}) = \frac{1}{R}\sum_{r=1}^{R} f(\Delta y_{t+k}|\mathbf{y_{1:t}}, \mathcal{F}^{(r)}) \xrightarrow{d}$$

$$\int_{\mathcal{F}} \overbrace{\prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta})}^{\text{Step 3}} \overbrace{f(\mathbf{z_{t+1:t+k}}|\mathbf{y_{1:t}}, \mathbf{z_{1:t}}, \boldsymbol{\theta})}^{\text{Step 2}} \overbrace{f(\mathbf{z_{1:t}}, \boldsymbol{\theta}|\mathbf{y_{1:t}})}^{\text{Step 1}} d\mathcal{F}.$$

Therefore, $\widehat{f}(\Delta y_{t+k}|\mathbf{y_{1:t}})$ can be constructed by averaging values of the predictive likelihood inside the summation above over R replications of our predictive simulator. In our forecasting exercise we set $R = 25000$ and discard the first 5000 burn-in draws. Importantly, since $\Delta y_{t+k}|\mathbf{y_{1:t}}, \mathcal{F}$ is normally distributed, one can readily evaluate $f(\Delta y_{t+k}|\mathbf{y_{1:t}}, \mathcal{F}^{(r)})$ to approximate $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$. Our predictive sampler can be summarized as follows:

At *every r-th* iteration of our MCMC algorithm and given data up to time t we (sequentially):
Step 1: simulate parameters and a $t \times 1$ vector of innovations based on the MCMC algorithm discussed in Section 2.5.
Step 2: simulate future innovations in $\mathbf{z_{t+1:t+k}}$ from the joint normal density, $f(\mathbf{z_{t+1:t+k}}|\mathbf{y_{1:t}}, \mathbf{z_{1:t}}, \boldsymbol{\theta})$, using variance parameters obtained in Step 1.
Step 3: use parameters and innovations produced in Steps 1 and 2 to generate draws for $\mathbf{\Delta y_{t+1:t+k}}$ from the joint normal density, $\prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta})$.

Once the algorithm above is performed R times, we have a proxy for the predictive density, $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$, which, in turn, enables us to construct metrics of point and density forecasts based on data up to point t. Next, we move forward one period and repeat Steps 1 through 3 to produce new (point and density) forecasts, only now with a sample size of $t+1$. We keep iterating in this fashion until point $T-1$.

Importantly, note that Step 3 can be carried out using a forecasting specification expressed in terms of innovations. For concreteness, consider again the CLARK-MSOE model. Taking first difference of the measurement equation for such model and moving it $t+s$ periods forward, for $s = 1, ..., k$, gives:

$$\Delta y_{t+s} = \mu_t + \sum_{j=1}^{s} \zeta_{t+s} + \eta_{t+s} + (\phi_1 - 1)c_{t+s-1} + \phi_2 c_{t+s-2} + \varepsilon_{t+s},$$

where $\varepsilon_{t+s} \sim \mathcal{N}\left(0,\ \sigma_\varepsilon^2\right)$. Therefore, the term $\sum_{j=1}^{s} \zeta_{t+s} + \eta_{t+s}$ can be obtained just as described in Step 2 and treated as predetermined in the expression above. Similarly, c_{t+s-j}, for $j = 1$ and 2, can be simulated using $\phi(L)c_{t+s-j} = \varepsilon_{t+s-j}$ and also be treated

as predetermined variables in the forecasting expression above. Lastly, μ_t can be recovered using the fact that for all $t = 1, ..., T$ (and for any given initial conditions), residual estimates of ε_t allow one to recover c_t (since $\phi(L)c_t = \varepsilon_t$), which yields τ_t (since $\tau_t = y_t - c_t$) and returns μ_t (since $\mu_t = \tau_t - (\tau_{t-1} + \eta_t)$).

Using simple algebraic manipulations, it is easy to see that an analogous strategy can be applied to all other UC models in order to obtain a predictive parameterization in terms of innovations. Also, for SSOE and RSOE variants, the in-sample correlation structure between states is preserved out-of-sample, hence, only innovations entering the model are simulated in Step 2 of our predictive sampler.

2.6.3 Results: Point Forecasts

To compare point forecast accuracy across models we compute the RMSFE for each k-step ahead prediction, y_{t+k}, defined as

$$\text{RMSFE} = \sqrt{\frac{\sum_{t=t_0}^{T-k} \left(y_{t+k}^o - \widehat{\mathbb{E}}(y_{t+k}|\mathbf{y_{1:t}}) \right)^2}{T - k - t_0 + 1}},$$

where y_{t+k}^o denotes the observed value of y_{t+k} that is known at time $t + k$, while t_0 and $T - k$ denote the first and last forecast generated respectively. Also, $\widehat{\mathbb{E}}(y_{t+k}|\mathbf{y_{1:t}})$ (i.e. an estimate of the mean of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$) is constructed by averaging over R draws from $\prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta})$ generated by the relevant innovations-based forecasting parameterization of an UC model, as discussed above.[26] Next, recall that we set t_0 =1971Q1 and T =2015Q2.

To make comparison of forecast performance across models easier, Table 2.3 reports RMSFE results relative to the local level-MSOE model. Specifically, entries in Table 2.3 denote the ratio of RMSFE between two competing models, where the RMSFE value in the denominator is always associated with the local level-MSOE model. Therefore, values less than one represent superior forecasting performance relative to the benchmark model. For simplicity, and when applicable, statistical significance results are reported only for models which outperform the local level MSOE model. Also, numbers in bold denote the best performing model for a specific forecast horizon.

Overall, results in Table 2.3 indicate that the orthogonality assumption seems suitable for short horizons. In particular the baseline local level-MSOE model is the best performing model for both CPI and PCE inflation at one, two and four quarter-

[26] Alternatively, one could use other statistics from the simulated $\widehat{f}(\Delta y_{t+k}|\mathbf{y_{1:t}})$ (e.g., median and mode) to construct the RMSFE. We have also produced RMSFE values based on the posterior median. Results are, however, broadly unchanged and available upon request.

ahead forecasts. When inflation is measured using the GDP deflator, the MNZ-MSOE model (i.e. with uncorrelated innovations) fares slightly better than the benchmark model at one and two-quarter ahead predictions. Albeit small, improvements are statistically significant at the 5% level. In contrast, allowing for correlation improves point forecast results at longer horizons for all measures of inflation we investigate. In particular, the local linear trend-RSOE model (i.e. when innovations to y_t and τ_t are perfectly correlated, but μ_t is still orthogonal) emerged as the best model at long run (point) forecasting. Such result is likely to reflect the usefulness of smoother measures of trend inflation (see Figure 2.3) in capturing long run inflation dynamics, as discussed in, e.g., Chan [2013] and Clark and Doh [2014]. Notably, best performing models generate improvements which are statistically significant at the 5% level relative to the orthogonal local level model. Also, I(2)-UC models which allow for some or all innovations to be correlated, such as the local linear trend-SSOE and CLARK-RSOE variants, outperform both the baseline model as well as their orthogonal counterparts at two, three and four year-ahead forecast horizons.

To summarize, results in Table 2.3 point to two main recommendations: (a) parsimonious UC models with orthogonal innovations seem appropriate for short-horizon point forecasts; and (b) for longer horizons, however, forecasting performance can be improved by relaxing the assumption of orthogonality between innovations. In particular, when looking exclusively at I(2)-UC models, RSOE and SSOE variants improve forecasting performance upon their orthogonal counterparts at longer horizons, regardless of the measure of price inflation used.

2.6.4 Results: Density Forecasts

As seen in Table 2.3, while the choice of the correlation structure affects point-forecast accuracy, differences induced by such different structures are in many cases small. This is, perhaps, unsurprising since point forecasts overlook the uncertainty surrounding such type of estimates. A simple way to illustrate this point is to think of two Gaussian predictive densities which display equivalent means but differ in terms of their variances. In an RMSFE sense, predictions from both densities are equivalent. On the other hand, when using forecast metrics which incorporate uncertainty around the prediction location the predictive accuracy between such densities would differ commensurate with their difference in variances.

Now, recall from our discussion in Section 2.3 that in a Bayesian setting there is a direct connection between the length (or volume) of the support of a predictive density and the amount of parameter uncertainty an UC model accommodates, as measured in terms of the restrictions an UC model specification imposes over its cor-

Table 2.3: Relative RMSFEs for US Quarterly Inflation Measures

CPI Inflation						
	Forecast Horizon					
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	1.030	1.048	1.040	1.024	1.024	1.035
Local Level-MSOE	**1.000**	**1.000**	**1.000**	1.000	1.000	1.000
MNZ-SSOE	1.030	1.051	1.040	1.023	1.026	1.036
MNZ-MSOE	1.007	1.002	1.001	1.001	1.001	0.998
MNZ-MSOE(UR)	1.055	1.089	1.066	1.042	1.045	1.061
Local Linear Trend-SSOE	1.121	1.008	1.021	0.995	0.979	0.945
Local Linear Trend-RSOE	1.205	1.058	1.021	**0.897***	**0.869***	**0.851***
Local Linear Trend-MSOE	1.164	1.035	1.052	1.036	1.023	0.991
CLARK-SSOE	1.781	1.457	1.307	1.211	1.021	0.977
CLARK-RSOE	1.167	1.032	1.028	0.936	0.906	0.878
CLARK-MSOE	1.401	1.321	1.551	1.301	0.998	0.995
GDP-Deflator Inflation						
	Forecast Horizon					
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	1.013	1.012	**0.991**	1.016	1.018	1.021
Local Level-MSOE	1.000	1.000	1.000	1.000	1.000	1.000
MNZ-SSOE	1.012	1.013	0.993	1.019	1.020	1.026
MNZ-MSOE	**0.997**	**0.998**	1.000	1.001	1.000	1.001
MNZ-MSOE(UR)	1.044	1.032	0.997	1.030	1.030	1.035
Local Linear Trend-SSOE	1.183	1.106	1.088	0.993	0.983	0.970
Local Linear Trend-RSOE	1.457	1.279	1.156	**0.977***	**0.956***	**0.950***
Local Linear Trend-MSOE	1.183	1.111	1.132	1.048	1.030	1.031
CLARK-SSOE	2.376	1.851	1.559	1.187	1.116	1.134
CLARK-RSOE	1.374	1.220	1.146	0.989	0.969	0.965
CLARK-MSOE	1.431	1.231	1.222	1.200	1.003	0.999
PCE-Deflator Inflation						
	Forecast Horizon					
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	1.017	1.030	1.023	1.020	1.022	1.026
Local Level-MSOE	**1.000**	**1.000**	**1.000**	1.000	1.000	1.000
MNZ-SSOE	1.020	1.030	1.024	1.016	1.018	1.025
MNZ-MSOE	1.003	1.003	1.004	1.002	1.008	0.999
MNZ-MSOE(UR)	1.041	1.059	1.047	1.036	1.040	1.051
Local Linear Trend-SSOE	1.179	1.063	1.037	0.994	0.990	0.970
Local Linear Trend-RSOE	1.283	1.106	1.021	**0.912***	**0.910***	**0.892***
Local Linear Trend-MSOE	1.196	1.082	1.061	1.029	1.025	1.006
CLARK-SSOE	2.098	1.538	1.331	1.106	1.104	1.038
CLARK-RSOE	1.242	1.079	1.025	0.933	0.931	0.908
CLARK-MSOE	1.401	1.321	1.551	1.301	0.998	0.995

* indicates superior forecast performance relative to the local level MSOE model at the 5% level of significance using a Diebold and Mariano [1995] test for equivalence in squared forecast errors.

responding reduced form ARIMA parameter space. Since changes in the correlation structure alter the parameter space over which $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$ is defined, forecast metrics which incorporate information within the full support of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$, are, thus, more likely to reflect stronger implications between out-of-sample performance and correlation structure.

A natural candidate to evaluate density forecasts is the sum of log predictive likelihoods:[27]

$$\sum_{t=t_0}^{T-k} \log \widehat{f}(\Delta y_{t+k}|\mathbf{y_{1:t}}).$$

For this metric, if the actual outcome $y^o_{t+k}|\mathbf{y_{1:t}}$ is unlikely under the density forecast, the value of $\log(f(\Delta y_{t+k} = \Delta y^o_{t+k}|\mathbf{y_{1:t}}))$ will be small and vise versa. Therefore, larger values of the sum of log predictive likelihoods indicate superior forecast performance. As before, for easy comparison, we present density forecast results relative to a baseline model given by the local level MSOE model. Specifically, entries in Table 2.4 denote the difference between the sum of log predictive likelihoods from a competing model relative to the sum of log predictive likelihoods from the baseline model. Therefore, positive numbers denote superior forecasting performance relative to the local level-MSOE model.

Overall, results in Table 2.4 reinforces the idea that modeling state correlation unambiguously affects density forecast performance at all horizons. In particular the Local level-SSOE is the best model for 1Q-ahead forecast, while the local linear trend-RSOE emerged as the best model for medium to longer forecasting horizons. As in the point forecast context, improvements for the best performing models are also statistically significant at the 5% level relative to the orthogonal local level baseline.

2.6.5 Correlation and Measures of Trend Inflation

In this section we report measures of trend CPI inflation (posterior median) for all models. All results presented in this section and the next are based on 250000 posterior draws after a burn-in step of 25000 using the MCMC algorithm described in Section 5 adjusted for each model accordingly.

Figure 2.3 shows that differences in trend inflation between MSOE and SSOE variants are minor. Trend inflation measures based on SSOE models are slightly more erratic than MSOE variants. Such result most likely reflects the fact that latent components in the SSOE case are recovered from an unique source of randomness

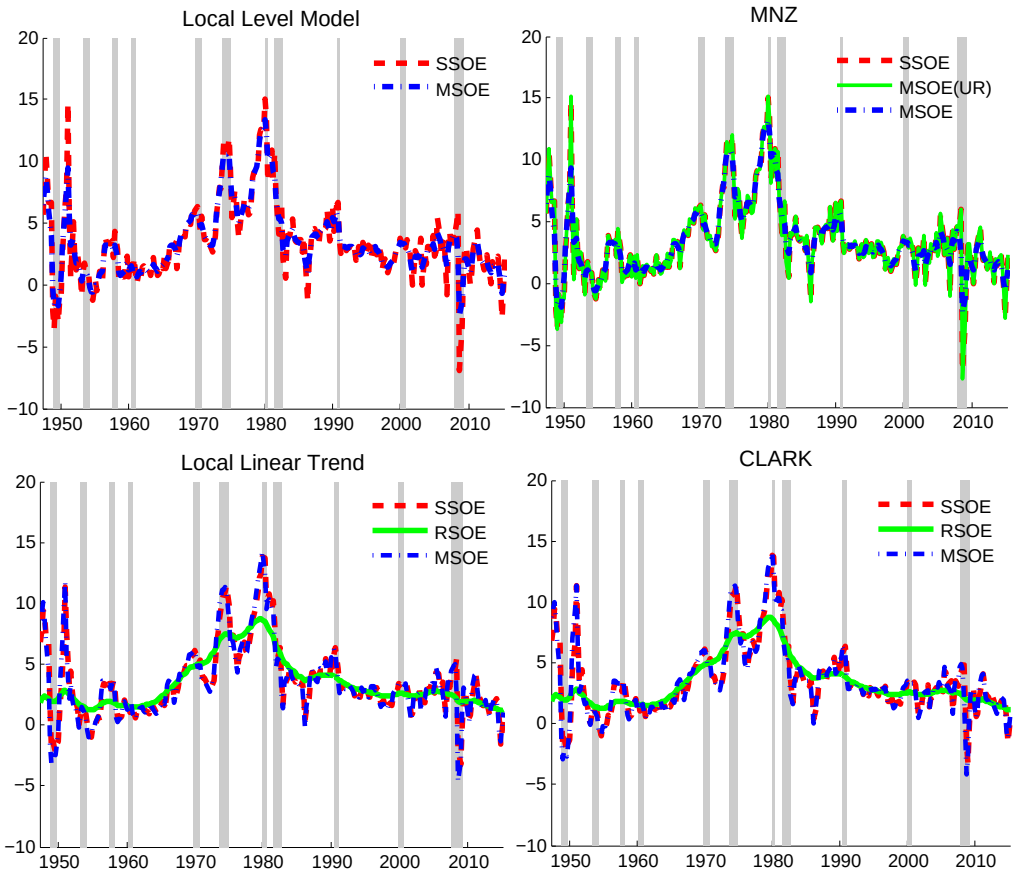
[27] See, e.g., Geweke and Amisano [2011] for a discussion on the predictive likelihood and its usefulness as a model comparison device.

Table 2.4: Sum of Log Predictive Likelihoods for US Quarterly Inflation Measures

CPI Inflation						
			Forecast Horizon			
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	**31.5***	29.1	26.6	19.1	16.4	17.8
Local Level-MSOE	0.0	0.0	0.0	0.0	0.0	0.0
MNZ-SSOE	7.7	5.1	9.3	9.2	6.1	4.3
MNZ-MSOE(UR)	13.6	16.1	18.7	18.1	20.3	11.1
MNZ-MSOE	-7.3	-10.2	-26.1	-13.1	9.1	5.2
Local Linear Trend-SSOE	14.3	12.1	13.2	10.4	9.7	15.6
Local Linear Trend-RSOE	11.7	13.5	**49.0***	**40.8***	**30.1***	**29.1***
Local Linear Trend-MSOE	-12.4	-10.1	-16.4	-20.4	-28.9	-41.4
CLARK-SSOE	27.1	**40.1***	41.1	39.1	27.4	26.1
CLARK-RSOE	29.1	39.8	24.9	29.6	38.8	39.1
CLARK-MSOE	-15.5	11.2	12.6	19.1	6.1	10.8
GDP-Implicit Price Deflator Inflation						
			Forecast Horizon			
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	**33.1***	24.2	23.1	21.9	19.1	19.8
Local Level-MSOE	0.0	0.0	0.0	0.0	0.0	0.0
MNZ-SSOE	7.4	7.2	8.1	8.3	7.7	5.1
MNZ-MSOE(UR)	14.1	17.2	15.9	18.5	19.3	11.7
MNZ-MSOE	-9.1	-13.6	-25.8	-18.2	3.1	2.0
Local Linear Trend-SSOE	14.3	14.7	11.9	20.1	17.4	13.1
Local Linear Trend-RSOE	10.1	18.9	28.5	**38.9***	**35.0***	**32.7***
Local Linear Trend-MSOE	-9.4	-12.4	-11.7	-16.5	-18.4	-30.8
CLARK-SSOE	29.1	**37.7***	**36.5***	34.1	29.1	24.9
CLARK-RSOE	22.5	29.1	25.4	21.5	30.3	29.5
CLARK-MSOE	-8.9	4.2	6.9	15.7	9.3	11.4
PCE-Implicit Price Deflator Inflation						
			Forecast Horizon			
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	**35.1***	25.3	23.1	29.4	19.1	20.9
Local Level-MSOE	0.0	0.0	0.0	0.0	0.0	0.0
MNZ-SSOE	6.1	6.5	10.9	11.2	8.3	9.1
MNZ-MSOE(UR)	11.3	18.2	15.7	18.7	21.9	20.2
MNZ-MSOE	-8.9	-11.1	-16.5	-17.3	8.2	7.1
Local Linear Trend-SSOE	11.4	19.9	**41.9***	34.1	37.1	33.1
Local Linear Trend-RSOE	19.1	23.1	34.5	**37.9***	**41.4***	**44.6***
Local Linear Trend-MSOE	-10.1	-18.9	-11.3	-12.1	-18.9	-36.4
CLARK-SSOE	17.9	**33.3***	36.1	35.3	29.1	30.6
CLARK-RSOE	17.9	16.7	22.1	26.1	37.4	38.9
CLARK-MSOE	-5.1	3.2	4.6	9.6	6.1	7.2

* indicates superior forecast performance relative to the local level MSOE model at the 5% level of significance using a Diebold and Mariano [1995] test for equivalence in squared forecast errors.

Figure 2.3: US Quarterly Measures of Annualized Trend (CPI) Inflation



Note: Shaded regions indicate recessions as recorded by the NBER

which encompasses all explained variability in inflation (or the price level). When comparing measures of trend inflation between MNZ-MSOE(UR) (which allows for imperfect correlation between c_t and τ_t) and its SSOE counterpart, differences are virtually imperceptible. One possible explanation for this result is presented in Figure 2.4, which shows that the implied posterior correlation between τ_t and c_t piles up near the (positive) perfect correlation region.

Another result is the fact that RSOE variants produce quite different measures of trend inflation relative to all other UC models. In particular, RSOE models produce measures of trend inflation which are quite smooth. As such, to the extent smoother measures of trend inflation are preferable for policy analysis, RSOE UC models can be perceived as more suitable correlation structures. From a statistical viewpoint (or in terms of the signal to noise ratio), such smoothness suggests that treating unex-

plained variability in the price level and in trend price level as purely measurement errors (which is essentially what RSOE variants do) provides a filtering strategy such that μ_t (i.e. trend inflation for I(2)-UC models) only reflects strong signals from price level changes.

2.6.6 Computation Efficiency

To assess the performance of an MCMC sampler, a common approach is to verify its mixing properties. In this sense, an MCMC algorithm with good mixing properties is one which allows the researcher to interpret parameter draws as independent realizations of a random variable. Consequently, an algorithm which produces strongly autocorrelated draws provides a clear indication of sampling inefficiency (or, equivalently, slow mixing).

In the context of state space models, the high dimensionality associated with the conditional posteriors in the state (or disturbance) smoothing steps can lead to poor mixing performance. Inefficient sampling can also be encountered when the conditional variance for any of the latent components is close to zero (see e.g. Frühwirth-Schnatter [2004] and Frühwirth-Schnatter and Wagner [2010]). The latter can be particularly an issue for I(2)-UC models since the conditional variance of y_t is decomposed across three stochastic processes, one of each, typically exhibiting very small conditional variance.

In practice, one alternative to improve MCMC simulation efficiency for UC models is to reparameterize the standard state-based representation given in Section 2.2.[28] Therefore, albeit UC models are reparameterized in this chapter with intent to address rank-reduction issues, a natural question is whether or not the disturbance-based parameterization combined with precision sampling techniques discussed in Section 2.5, lead to an MCMC sampler with good mixing properties. To address such questions we report inefficiency factors of the posterior draws for all parameters and innovations using a common metric (see, e.g. Chib [2001]) given by:

$$1 + 2\sum_{j=1}^{J} \rho_j,$$

where ρ_j is the sample autocorrelation at lag j through lag J. In our empirical application we set J to be large enough until autocorrelation tapers off. Clearly, in an ideal setting where MCMC draws are virtually independent draws, inefficiency factors

[28] For example, Papaspiliopoulos et al. [2003], Frühwirth-Schnatter [2004] and Frühwirth-Schnatter and Wagner [2010] discuss how MCMC sampling efficiency improvements can be achieved by parameterizing UC models with variance parameters in the measurement rather than state equations.

Figure 2.4: Density Estimate for the Correlation Between τ_t and c_t Under MNZ-MSOE(UR) Model

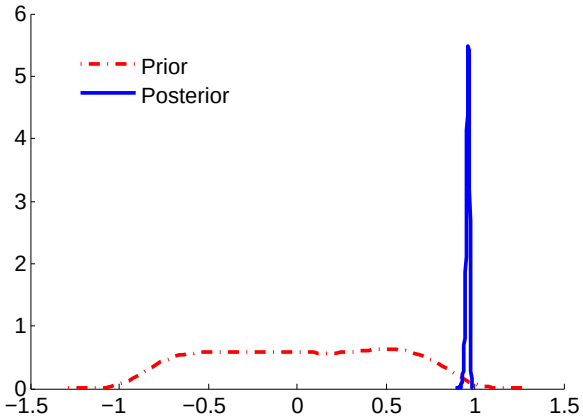
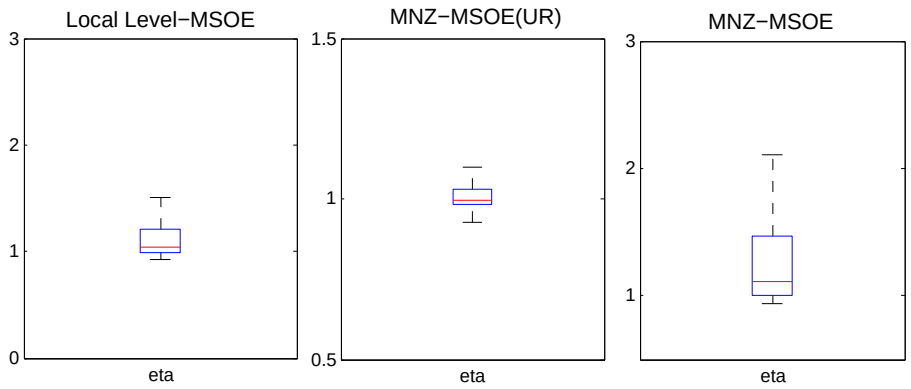


Figure 2.5: Inefficiency Factors for Disturbance Smoothing (I(1)-UC Models)



should be 1. As a rule of thumb, inefficiency factors around 20 based on the metric above are typically interpreted as an indication of fast mixing.[29] Table 2.5 reports the inefficiency factors for parameter draws associated with each of the eleven models. Figure 2.5 and Figure 2.6 report inefficiency factors for the $T \times 1$ disturbance vectors $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ for I(1) and I(2)-UC models, respectively, when disturbance smoothing is required (i.e. RSOE and MSOE variants). Notably, for disturbance smoothing, instead of reporting inefficiency factors for each one of the T elements in $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$, we follow Chan [2015] and use boxplots to present the information on inefficiency factor visually. In particular, the middle line denotes the median inefficiency factor based on a sample T inefficiency factors constructed for each element in $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$. Similarly, lower and upper lines respectively represent the 25 and 75 percentiles, while whiskers extend to the maximum and minimum. All and in all, our results below suggest that the algorithm developed in Section 2.5 is quite efficient in terms of generating parameters and innovations draws which are not strongly autocorrelated. In addition to inefficiency factors, other metrics of interest to assess MCMC sampling performance are the acceptance ratio on the proposal density in the Metropolis-Hastings step as well as computational speed of our algorithm. Acceptance rates range from 70% to 90% depending on the model, thus, suggesting that the Gaussian proposal well approximates the conditional posterior density of $\boldsymbol{\theta}_{-\sigma^2}|\widetilde{\mathbf{y}}, \mathbf{z}, \boldsymbol{\sigma^2}$. In terms of computational speed, estimation of all models where 10000 draws are sampled for each model takes less than 180 seconds.

[29] Another way to interpret the inefficiency factor adopted here is to think that an inefficiency factor of 100 means that approximately 10000 posterior draws are required to convey the same information as 100 independent draws.

Figure 2.6: Inefficiency Factors for Disturbance Smoothing (I(2)-UC Models)

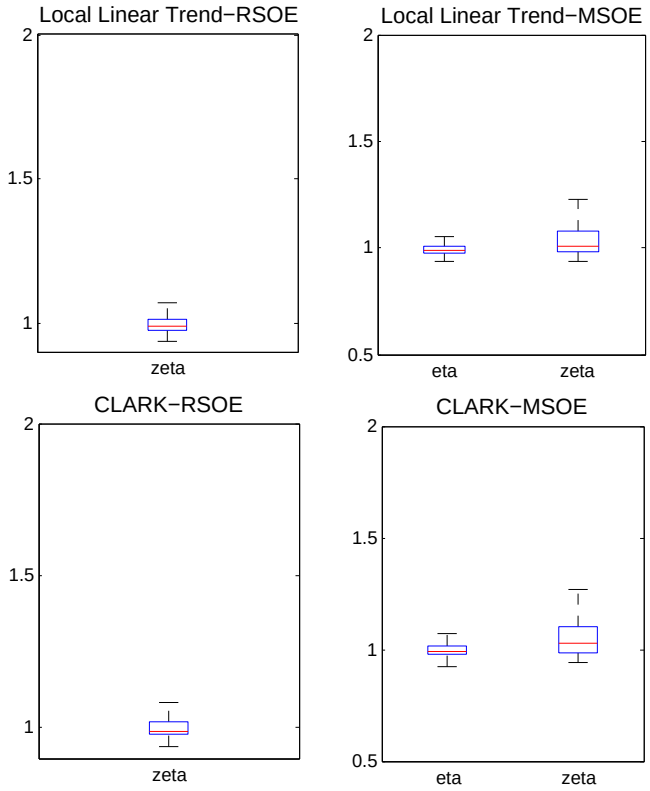


Table 2.5: Inefficiency Factors for Parameter Sampling

Model	σ_ε^2	σ_η^2	σ_ζ^2	κ_τ	κ_μ	ϕ_1	ϕ_2
Local Level-SSOE	1.99	–	–	1.27	–	–	–
Local Level-MSOE	7.33	10.17	–	–	–	–	–
MNZ-SSOE	2.11	–	–	2.05	–	1.73	1.22
MNZ-MSOE(UR)	2.49	17.15	–	10.06	–	4.56	5.21
MNZ-MSOE	8.21	13.23	–	–	–	8.41	5.27
Local Linear Trend-SSOE	3.11	–	–	3.10	13.17	–	–
Local Linear Trend-RSOE	1.77	–	11.32	6.41	–	–	–
Local Linear Trend-MSOE	9.91	18.41	3.82	–	–	–	–
CLARK-SSOE	1.85	–	–	5.19	6.01	3.01	1.73
CLARK-RSOE	3.14	–	13.34	5.08	–	1.45	2.10
CLARK-MSOE	7.11	1.99	20.77	–	–	6.10	6.31

2.7 Concluding Remarks and Extensions

In this chapter we have studied the relationship between state correlation and out-of-sample performance within a Bayesian framework. Given the recent interest in the literature on UC models to forecast inflation, we focused on such class of state space model. Akin to Stock and Watson [2007], in our empirical application we used inflation measures based on the CPI and the real GDP and PCE price deflators.

Following a substantial forecasting exercise, we demonstrated that modeling state correlation has relevant effects to forecasting performance. Specifically, allowing for correlated state variables generated statistically significant improvements in both point and density forecasts relative to the usual orthogonal UC model counterparts. In particular, a new approach to model state correlation which combines features from orthogonal as well as perfectly correlated states emerged as one of the best performing models in terms of both point and density forecasts. Such variant also generates smooth measures of trend inflation, which is typically a desirable feature for policy analysis.

Another contribution from this chapter was to develop a new algorithm based on precision sampling techniques and properties of Toeplitz matrices to conduct fast MCMC simulation of UC models with full and reduced rank covariance matrices. In our study, rank reduction stemmed from allowing for perfect correlation between two or more states. For future research, it would be interesting to extend the models and algorithms developed here to incorporate other important features for forecasting such as stochastic volatility as well as formulate multivariate versions of the UC models entertained in this study.

Appendix 2.A Appendix

2.A.1 Proof of Proposition 2.3.1

To prove part 3.1(a), recall first that the Local Linear Trend-MSOE model in Section 2.2.2 can be represented as a sum of three MA processes:

$$\Delta^2 y_t = \zeta_t + \eta_t - \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}.$$

To obtain the RSOE variant we set: $\eta_t = \kappa_\tau \varepsilon_t$. Therefore the expression above can be recast as:

$$\Delta^2 y_t = \zeta_t + \left[\kappa_\tau(1-L) + 1 - 2L + L^2\right]\varepsilon_t. \tag{2.A.1}$$

The existence of a reduced form ARIMA(0,2,2) representation follows from Granger's lemma (see Granger and Newbold [1986], p. 28-30). To be precise, note that the expression in the right-hand-side of (2.A.1) denotes the sum of two independent MA processes, namely a MA(0) (i.e. ζ_t) and a MA(2) (i.e. $\left[\kappa_\tau(1-L) + 1 - 2L + L^2\right]\varepsilon_t$) process. Granger's lemma, thus, ensures that the resulting process will be an MA(q) polynomial such that $q = \max(0, 2) = 2$.

Next, to show that the resulting reduced form ARIMA specification is invertible, we apply Corollary 1.2. to Theorem 1 in Teräsvirta [1977], which states that the sum of seemingly unrelated MA polynomials is invertible if at least one of the polynomials is a white noise process. The latter is satisfied by the assumption: $\zeta \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_\zeta)$.

The proof of claim 3.1(b) follows an analogous strategy. Akin to the local linear trend model, straightforward algebraic manipulations to the CLARK-MSOE model in Section 2.2.2, yield an equivalent representation in terms of MA polynomials:

$$\phi(L)\Delta^2 y_t = \phi(L)\zeta_t + \phi(L)(1-L)\eta_t + (1-L)^2\varepsilon_t.$$

Using the fact that $\eta_t = \kappa_\tau \varepsilon_t$ gives us:

$$\phi(L)\Delta^2 y_t = \phi(L)\zeta_t + \left[\kappa_\tau\phi(L)(1-L) + (1-L)^2\right]\varepsilon_t.$$

By virtue of Granger's lemma, the right-hand side in the expression above has a MA(3) reduced form representation. To ensure the resulting MA polynomial is invertible, we use use Theorem 1 in Teräsvirta [1977], which states that the sum of two or more (possibly correlated) MA polynomials is invertible if and only if such polynomials do not share common roots of modulus one. The latter is satisfied by noting that the stationarity condition of the AR polynomial precludes $\phi(\pm 1) = 0$.

Therefore, as long as c_t is a stationary state, the MA polynomials above, $\phi(L)\zeta_t$ and $\left[\kappa_\tau \phi(L)(1-L) + (1-L)^2\right]\varepsilon_t$, do not share such type of roots.

2.A.2 Proof of Proposition 2.4.1

First, note that simple application of Bayes' rule (and omitting initial conditions to make notation less cumbersome) yields:

$$f(\mathbf{z}|\mathbf{y},\boldsymbol{\theta}) = \frac{f(\mathbf{y},\mathbf{z}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta})}$$
$$f(\boldsymbol{\theta}|\mathbf{y},\mathbf{z}) = \frac{f(\mathbf{y},\mathbf{z}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y},\mathbf{z})},$$

where $f(\mathbf{y},\mathbf{z}|\boldsymbol{\theta})$ denotes the *complete-data* likelihood function. Clearly, $f(\mathbf{y},\mathbf{z}|\boldsymbol{\theta})$ enters the kernel of both conditional posterior densities above. Thus, showing that the covariance matrix associated with the *complete-data* likelihood function is not invertible validates the claim in Proposition 2.4.1. To do so, we prove Proposition 2.4.1 by contradiction. Since all models entertained in this chapter are linear Gaussian, let $\mathbf{y},\mathbf{z}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{a},\ \mathbf{A})$ and assume $\mathbf{A}$ denotes a nonsingular covariance matrix. Therefore, using standard results in matrix algebra (see, e.g., Anderson [1984]) we partition $\mathbf{A}$, such that:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{pmatrix}$$

where $\mathbf{A_{11}}$ and $\mathbf{A_{22}}$ are symmetric submatrices denoting covariance matrices corresponding to some partition of the elements in $\mathbf{z}$. Moreover, matrix inversion results (see, e.g. Theorem A.3.3 in Anderson [1984]), yield:

$$\mathbf{A}^{-1} = \begin{pmatrix} (\mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{21}})^{-1} & -(\mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{21}})^{-1}\mathbf{A_{12}}\mathbf{A_{22}}^{-1} \\ -\mathbf{A_{22}}^{-1}\mathbf{A_{12}}(\mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{21}})^{-1} & \mathbf{A_{22}}^{-1}\mathbf{A_{21}}(\mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{21}})^{-1}\mathbf{A_{12}}\mathbf{A_{22}}^{-1} + \mathbf{A_{22}}^{-1} \end{pmatrix}.$$

Now, recall that $\mathbf{z} = \{\boldsymbol{\tau}, \boldsymbol{\mu}\}$. Since $\mathbf{y}$, $\boldsymbol{\tau}$ and $\boldsymbol{\mu}$ are exchangeable random vectors, then a feasible partition of $\mathbf{A}$ describing the same joint density, $\mathbf{y},\boldsymbol{\tau},\boldsymbol{\mu}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{a},\ \mathbf{A})$, is one which assigns the joint covariance matrix of $\boldsymbol{\mu}|\boldsymbol{\theta}$ to $\mathbf{A_{11}}$ and the joint covariance matrix of $\mathbf{y},\boldsymbol{\tau}|\boldsymbol{\theta}$ to $\mathbf{A_{22}}$. To complete the proof, it suffices to remember that perfect correlation between $\mathbf{y}|\boldsymbol{\theta}$ and $\boldsymbol{\tau}|\boldsymbol{\theta}$, which occurs in both SSOE and RSOE variants, introduces singularity to $\mathbf{A_{22}}$ through row and column-wise linear dependence, which in turn makes $\mathbf{A}$ non-invertible (since $\mathbf{A_{22}}^{-1}$ cannot be constructed), thus, contradicting $\mathbf{A}$ being nonsingular. Since $\mathbf{A}$ cannot be inverted, it is rank-deficient.

2.A.3 Derivation of the Restrictions Over the MA Parameter Space Shown in Figure 2.1

In this section we show how to obtain the nonlinear restrictions over the MA parameter space as show in Figure 2.1. To that end, for the MSOE and RSOE variants we take a similar approach as in, e.g., Watson [1986] and Harvey [1989], thereby comparing the autocorrelation function generated by the local linear trend model with its corresponding (unrestricted) reduced form counterpart. Remember, underpinning such comparison is the fact that UC models can also be perceived as structural representations of an ARIMA model. Hence, let $\rho(s)$ denote the autocorrelation function associated with the reduced form ARIMA (0,2,2) in (2.6) and its corresponding representation given in (2.5), i.e., in terms of the local linear trend model parameters. Thus, we have:

$$\rho(s) = \begin{cases} \frac{\varphi_1(1+\varphi_2)}{1+\varphi_1^2+\varphi_2^2} = -\frac{\sigma_\eta^2+4\sigma_\varepsilon^2+4\sigma_{\eta\varepsilon}+2\sigma_{\zeta\varepsilon}+\sigma_{\eta\zeta}}{2\sigma_\eta^2+6\sigma_\varepsilon^2+\sigma_\zeta^2+2\sigma_{\eta\zeta}+6\sigma_{\eta\varepsilon}+2\sigma_{\zeta\varepsilon}} & \text{if } s=1, \\ \frac{\varphi_2}{1+\varphi_1^2+\varphi_2^2} = \frac{\sigma_\varepsilon^2+\sigma_{\eta\varepsilon}+\sigma_{\zeta\varepsilon}}{2\sigma_\eta^2+6\sigma_\varepsilon^2+\sigma_\zeta^2+2\sigma_{\eta\zeta}+6\sigma_{\eta\varepsilon}+2\sigma_{\zeta\varepsilon}} & \text{if } s=2, \\ 0 \quad \text{if } s \geq 3, \end{cases} \tag{2.A.2}$$

where $\sigma_{i,j} = \mathbb{C}\text{ov}(i,j)$, for $i \neq j$ and $i,j = \eta_t, \zeta_t,$ and ε_t. Next, we apply the correlation restrictions implied by the Local Linear Trend-MSOE and RSOE variants to the system above to derive the restrictions (over the MA space) associated with these two models.

• *Local Linear Trend-MSOE*

Setting $\sigma_{\eta\varepsilon} = \sigma_{\eta\zeta} = \sigma_{\zeta\varepsilon} = 0$ in the 1^{st} and 2^{nd} autocorrelations expressions in (2.A.2) and rearranging terms yields:

$$\varphi_1 = -(4+q)\frac{\varphi_2}{1+\varphi_2}, \tag{2.A.3a}$$

$$\varphi_2 = g_1\sigma_\varepsilon^2. \tag{2.A.3b}$$

Where $q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} > 0$ and $g_1 = \frac{1+\varphi_1^2+\varphi_2^2}{2\sigma_\eta^2+6\sigma_\varepsilon^2+\sigma_\zeta^2} > 0$. Using these two positive constraints, then, from (2.A.3b) it is easy verify that the Local Linear Trend-MSOE model can only generate values of φ_2 such that $\varphi_2 > 0$. Next, using $\varphi_2 > 0$ and $-(4+q) < 0$ in (2.A.3a) yields $\varphi_1 < 0$. Therefore, the Local Linear Trend-MSOE model is compatible with reduced form MA parameters located only in the $\varphi_1 < 0$ and $\varphi_2 > 0$ quadrant. To pin down the exact restrictions, note that equation (2.A.3a) describes a hyperbola in the (φ_1, φ_2) space with eccentricity (i.e. degree of flatness) controlled by $-(4+q) \in$

$(-\infty, -4)$. Consequently, one can numerically evaluate (2.A.3a) for a wide range of values of $-(4+q) \in (-\infty, -4)$ to construct the space denoted by all such hyperbolas which intersect with the $\varphi_1 < 0$ and $\varphi_2 > 0$ quadrant and the the invertibility space (i.e., the triangular region). The resulting region is the admissibility region of MA parameters for the Local Linear Trend-MSOE model as shown in the bottom panel in Figure 2.1.

• *Local Linear Trend-RSOE*

First, recall that Local Linear Trend-RSOE model imposes the following: $\eta_t = \kappa_\tau \varepsilon_t; \sigma_{\eta\varepsilon} = \kappa_\tau \sigma_\varepsilon^2$; $\sigma_\eta^2 = \kappa_\tau^2 \sigma_\varepsilon^2$ and $\sigma_{\eta\zeta} = \sigma_{\zeta\varepsilon} = 0$. Plugging these into (2.A.2) and rearranging terms yields:

$$\varphi_1 = -\frac{(2+\kappa_\tau)^2}{1+\kappa_\tau}\frac{\varphi_2}{1+\varphi_2}, \tag{2.A.4a}$$

$$\varphi_2 = g_2 (1+\kappa_\tau)\sigma_\varepsilon^2, \tag{2.A.4b}$$

where $g_2 = \frac{1+\varphi_1^2+\varphi_2^2}{(2\kappa_\tau^2+6+6\kappa_\tau)\sigma_\varepsilon^2+\sigma_\zeta^2} > 0$ and $\kappa_\tau \in \mathbb{R}$ such that $\kappa_\tau \neq -1$. Since the loading parameter, κ_τ, can take both negative and positive values, from (2.A.4b) one can see that (unlike the MSOE variant) φ_2 can now take both negative and positive values. More precisely, it is easy to check that: $\varphi_2 < 0$ (> 0) if $\kappa_\tau < -1$ (> -1). Next, from (2.A.4a), again, we have a hyperbola in the (φ_1, φ_2) space. Noting that the eccentricity of such hyperbola is now controlled by $-\frac{(2+\kappa_\tau)^2}{1+\kappa_\tau}$, such that $-\frac{(2+\kappa_\tau)^2}{1+\kappa_\tau} \in \mathbb{R}^+$ if $\kappa_\tau < -1$ and $-\frac{(2+\kappa_\tau)^2}{1+\kappa_\tau} \leq -4$ if $\kappa_\tau > -1$, numerical evaluation of such function for a wide range of eccentricity values generates the compatibility region presented in the center panel of Figure 2.1. In particular, the negative constraint to φ_1 is preserved, since positive values of φ_2 coincide with negative values of $-\frac{(2+\kappa_\tau)^2}{1+\kappa_\tau}$ and vice-versa.

• *Local Linear Trend-SSOE*

SSOE models provide a direct mapping between UC and reduced form ARIMA parameters. We explore this fact to readily derive the parameter space restrictions for the Local Linear Trend-SSOE case. Recall first that for such model we set: $\eta_t = \kappa_\tau \varepsilon_t$ and $\zeta_t = \kappa_\mu \varepsilon_t$. Plugging these into (2.5) yields:

$$\begin{aligned}
\Delta^2 y_t &= (\kappa_\tau \varepsilon_t + \varepsilon_t) + \left(\kappa_\mu \varepsilon_{t-1} - \varepsilon_{t-1} - 2\varepsilon_{t-1}\right) + \varepsilon_{t-2} \\
\Delta^2 y_t &= (1+\kappa_\tau)\varepsilon_t + \left(\kappa_\mu - 3\right)\varepsilon_{t-1} + \varepsilon_{t-2} \\
\Delta^2 y_t &= u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2},
\end{aligned}$$

such that $u_t = (1+\kappa_\tau)\,\varepsilon_t$, i.e., $u_t \sim \mathcal{N}(0, (1+\kappa_\tau)^2\,\sigma_\varepsilon^2)$; $\varphi_1 = \frac{(\kappa_\mu - 3)}{(1+\kappa_\tau)}$ and $\varphi_2 = \frac{1}{(1+\kappa_\tau)}$. In other words, the Local Linear Trend SSOE model allows one to back out MA coefficients by simply using estimates of the loading parameters, κ_τ and κ_μ. In particular, it is easy to verify that there is a one-to-one mapping between φ_2 and κ_τ as well as between φ_1 and κ_μ (for any given value of κ_τ). Consequently, the SSOE variant opens the (φ_1, φ_2) parameter space relative to the restrictions in the RSOE and MSOE variants. Specifically, as shown in the top panel of Figure 2.1, the Local Linear Trend SSOE model is compatible with the full invertibility space, except for the horizontal axis, since $\varphi_2 = \frac{1}{(1+\kappa_\tau)}$ is not defined at zero.

2.A.4 Invertibility and Stationarity for SSOE Models

To derive the invertibility and stationarity conditions for SSOE models, once again, we explore the explicit mapping that exists between such models and their ARIMA representation.

•*Local Level-SSOE*

Taking first difference of the measurement equation of the Local Level-SSOE model yields:

$$\begin{aligned}\Delta y_t &= (1+\kappa_\tau)\varepsilon_t - \varepsilon_{t-1},\\ \Delta y_t &= u_t + \varphi_1 u_{t-1},\end{aligned}$$

where $u_t = (1+\kappa_\tau)\varepsilon_t$ and $\varphi_1 = -\frac{1}{1+\kappa_\tau}$. Therefore, to ensure invertibility we keep draws of κ_τ that satisfy $\left|-\frac{1}{1+\kappa_\tau}\right| < 1$.

•*Local Linear Trend-SSOE*

Taking second differences of the measurement equation of the Local Linear Trend-SSOE model yields:

$$\begin{aligned}\Delta^2 y_t &= (1+\kappa_\tau+\kappa_\mu)\varepsilon_t - (\kappa_\tau+2)\varepsilon_{t-1} + \varepsilon_{t-2},\\ \Delta^2 y_t &= u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2},\end{aligned}$$

where $u_t = (1+\kappa_\tau+\kappa_\mu)\varepsilon_t$, $\varphi_1 = -\frac{\kappa_\tau+2}{1+\kappa_\tau+\kappa_\mu}$ and $\varphi_2 = \frac{1}{1+\kappa_\tau+\kappa_\mu}$. Therefore, noting that the Local Linear Trend-SSOE model is observationally equivalent to a reduced-form MA(2) process implies that invertibility of the former is ensured by keeping draws

of κ_τ and κ_μ such that:

$$\begin{cases} -\frac{\kappa_\tau+2}{1+\kappa_\tau+\kappa_\mu} + \frac{1}{1+\kappa_\tau+\kappa_\mu} & < 1, \\ -\frac{\kappa_\tau+2}{1+\kappa_\tau+\kappa_\mu} - \frac{1}{1+\kappa_\tau+\kappa_\mu} & < 1, \\ \left|\frac{1}{1+\kappa_\tau+\kappa_\mu}\right| < 1, \end{cases}$$

in other words, the standard invertibility conditions for a reduced form MA(2) process.

•*MNZ-SSOE*

Recall that we set $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$. Taking first difference of the measurement equation for the MNZ-SSOE model and rearranging terms yields:

$$\phi(L)\Delta y_t = (1+\kappa_\tau)\varepsilon_t - (\kappa_\tau\phi_1 + 1)\varepsilon_{t-1} - \kappa_\tau\phi_2\varepsilon_{t-2}, \tag{2.A.5}$$

$$\phi(L)\Delta y_t = u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2}, \tag{2.A.6}$$

where $u_t = (1+\kappa_\tau)\varepsilon_t$, $\varphi_1 = -\frac{\kappa_\tau\phi_1+1}{1+\kappa_\tau}$ and $\varphi_2 = -\frac{\kappa_\tau\phi_2}{1+\kappa_\tau}$. Therefore, invertibility is ensured by keeping draws of κ_τ, ϕ_1 and ϕ_2, such that:

$$\begin{cases} -\frac{\kappa_\tau\phi_1+1}{1+\kappa_\tau} - \frac{\kappa_\tau\phi_2}{1+\kappa_\tau} & < 1, \\ -\frac{\kappa_\tau+2}{1+\kappa_\tau+\kappa_\mu} + \frac{\kappa_\tau\phi_2}{1+\kappa_\tau} & < 1, \\ \left|-\frac{\kappa_\tau\phi_2}{1+\kappa_\tau}\right| < 1. \end{cases}$$

To ensure stationarity of $\phi(L)$ note that the AR part in (2.A.5) and (2.A.6) are the same. Hence, keeping draws of ϕ_1 and ϕ_2 such that $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$, ensures stationarity of the AR(2) polynomial, $\phi(L)$.

•*CLARK-SSOE*

Taking second differences of the measurement equation of the CLARK-SSOE model yields:

$$\begin{aligned} \phi(L)\Delta^2 y_t &= (1+\kappa_\tau+\kappa_\mu)\varepsilon_t - (\kappa_\tau\phi_1 + \kappa_\mu\phi_1 + \kappa_\tau + 2)\varepsilon_{t-1} + \\ &\quad + (-\kappa_\tau\phi_2 - \kappa_\mu\phi_2 + \kappa_\tau\phi_1 + 1)\varepsilon_{t-2} + \kappa_\tau\phi_2\varepsilon_{t-3}, \\ \phi(L)\Delta^2 y_t &= u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \varphi_3 u_{t-3}, \end{aligned}$$

where $u_t = (1+\kappa_\tau+\kappa_\mu)\varepsilon_t$, $\varphi_1 = -\frac{(\kappa_\tau\phi_1+\kappa_\mu\phi_1+\kappa_\tau+2)}{1+\kappa_\tau+\kappa_\mu}$, $\varphi_2 = -\frac{\kappa_\tau\phi_2-\kappa_\mu\phi_2+\kappa_\tau\phi_1+1}{1+\kappa_\tau+\kappa_\mu}$ and $\varphi_3 = \frac{\kappa_\tau\phi_2}{1+\kappa_\tau+\kappa_\mu}$.

Therefore, invertibility is ensured by keeping draws of κ_τ, κ_μ, ϕ_1 and ϕ_2 such that the roots of the MA polynomial, $1 + \phi_1 L + \phi_2 L + \phi_3 L^3$, lie outside the unit circle. Also, note that, as in the MNZ-SSOE case, the AR polynomial for ARIMA(2,2,3) process above is the same regardless if it is presented in terms of ε_t (i.e. the SSOE representation of the ARIMA model) or u_t. Hence, keeping draws of ϕ_1 and ϕ_2 such that $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$, ensures stationarity of the AR(2) polynomial, $\phi(L)$.

2.A.5 Disturbance-Based Parameterization for I(1) and I(2)-UC Models

In this section we show, as claimed in Section 2.5, that all eleven UC models entertained in this chapter can be recast into a general framework where innovations are moved to the measurement equation and state equations become white noise processes. Importantly, such a general parameterization allows one to estimate all models using the algorithm developed in Section 2.5 for appropriately defined expressions underlying reduced form matrices and vectors (see below).

In what follows, we adopt a derivation strategy similar to the one presented in (2.12a)-(2.12d), i.e., we apply the commutative property of lower triangular Toeplitz matrices to construct a disturbance-based parameterization which is conducive to quick MCMC estimation. In the interest of brevity, we do not repeat the algebra for the CLARK-MSOE model here. Also, note that the derivation starting point for I(2)-UC models is equation (2.12b). For I(1)-UC models, a slightly modified version of (2.12b) is applied:

$$\mathbf{Hy} = \boldsymbol{\iota}_0 \tau_0 + \boldsymbol{\eta} + \mathbf{Hc}.$$

It is easy to check that the latter can be seen as the (matrix notation of the) measurement equation for all I(1)-UC models with both sides pre-multiplied by $\mathbf{H}$ and prior to any state correlation adjustments.

With these ideas in mind, adjusting each UC model according to its correlation structure, as discussed in Section 2.2, is ensued by straightforward algebraic manipulations.

I(2)-UC Models

•*CLARK-RSOE*

$$
\begin{aligned}
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \boldsymbol{\zeta} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HHc} \\
\mathbf{H}_\phi\mathbf{HHy} &= \mathbf{H}_\phi\mathbf{H}\boldsymbol{\iota}_0\tau_0 + \mathbf{H}_\phi\boldsymbol{\iota}_0\mu_0 + \mathbf{H}_\phi\boldsymbol{\zeta} + \left(\kappa_\tau\mathbf{H}_\phi\mathbf{H} + \mathbf{HH}\right)\boldsymbol{\varepsilon} \\
\mathbf{H}_\phi\mathbf{HH}(\mathbf{A}\setminus\mathbf{y}) &= \left(\mathbf{H}_\phi\mathbf{H}(\mathbf{A}\setminus\boldsymbol{\iota}_0)\;\; \mathbf{H}_\phi(\mathbf{A}\setminus\boldsymbol{\iota}_0)\right)(\tau_0\;\mu_0)' + \mathbf{H}_\phi(\mathbf{A}\setminus\boldsymbol{\zeta}) + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}
$$

$$
\text{s.t.}\begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{H}_\phi\mathbf{H} + \mathbf{HH}, \\
\mathbf{X_0} &= \left(\mathbf{H}_\phi\mathbf{H}(\mathbf{A}\setminus\boldsymbol{\iota}_0)\;\; \mathbf{H}_\phi(\mathbf{A}\setminus\boldsymbol{\iota}_0)\right), \\
\mathbf{X_1} &= \mathbf{H}_\phi, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= (\tau_0\;\mu_0)', \\
\widetilde{\mathbf{y}} &= \mathbf{H}_\phi\mathbf{HH}(\mathbf{A}\setminus\mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{A}\setminus\boldsymbol{\zeta}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}
$$

•*CLARK-SSOE*

$$
\begin{aligned}
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \kappa_\mu\boldsymbol{\varepsilon} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HHc} \\
\mathbf{H}_\phi\mathbf{HHy} &= \mathbf{H}_\phi\mathbf{H}\boldsymbol{\iota}_0\tau_0 + \mathbf{H}_\phi\boldsymbol{\iota}_0\mu_0 + \left(\kappa_\mu\mathbf{H}_\phi + \kappa_\tau\mathbf{H}_\phi\mathbf{H} + \mathbf{HH}\right)\boldsymbol{\varepsilon} \\
\mathbf{H}_\phi\mathbf{HH}(\mathbf{A}\setminus\mathbf{y}) &= \left(\mathbf{H}_\phi\mathbf{H}(\mathbf{A}\setminus\boldsymbol{\iota}_0)\quad \mathbf{H}_\phi(\mathbf{A}\setminus\boldsymbol{\iota}_0)\right)(\tau_0\;\mu_0)' + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}
$$

$$
\text{s.t.}\begin{cases}
\mathbf{A} &= \kappa_\mu\mathbf{H}_\phi + \kappa_\tau\mathbf{H}_\phi\mathbf{H} + \mathbf{HH}, \\
\mathbf{X_0} &= \left(\mathbf{H}_\phi\mathbf{H}(\mathbf{A}\setminus\boldsymbol{\iota}_0)\;\; \mathbf{H}_\phi(\mathbf{A}\setminus\boldsymbol{\iota}_0)\right), \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= (\tau_0\;\mu_0)', \\
\widetilde{\mathbf{y}} &= \mathbf{H}_\phi\mathbf{HH}(\mathbf{A}\setminus\mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}
$$

•*Local Linear Trend-MSOE*

$$\begin{aligned}
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \boldsymbol{\zeta} + \mathbf{H}\boldsymbol{\eta} + \mathbf{HH}\boldsymbol{\varepsilon} \\
\mathbf{y} &= (\mathbf{H} \setminus \boldsymbol{\iota}_0 \;\; (\mathbf{HH}) \setminus \boldsymbol{\iota}_0)\,(\tau_0 \;\mu_0)' + (\mathbf{HH})\backslash\boldsymbol{\zeta} + \mathbf{H}\backslash\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases}
\mathbf{A} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_0} &= (\mathbf{H} \setminus \boldsymbol{\iota}_0 \;\; (\mathbf{HH}) \setminus \boldsymbol{\iota}_0)\,, \\
\mathbf{X_1} &= \mathbf{I}_{(T\times T)}, \\
\mathbf{X_2} &= \mathbf{I}_{(T\times T)}, \\
\mathbf{z_0} &= (\tau_0 \;\mu_0)', \\
\widetilde{\mathbf{y}} &= \mathbf{y}, \\
\widetilde{\boldsymbol{\zeta}} &= (\mathbf{HH})\backslash\boldsymbol{\zeta}, \\
\widetilde{\boldsymbol{\eta}} &= (\mathbf{H})\backslash\boldsymbol{\eta}.
\end{cases}$$

•*Local Linear Trend-RSOE*

$$\begin{aligned}
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \boldsymbol{\zeta} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HH}\boldsymbol{\varepsilon} \\
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \boldsymbol{\zeta} + (\kappa_\tau\mathbf{H} + \mathbf{HH})\,\boldsymbol{\varepsilon} \\
\mathbf{HH}(\mathbf{A} \setminus \mathbf{y}) &= (\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota}_0) \;\; (\mathbf{A} \setminus \boldsymbol{\iota}_0))\,(\tau_0 \;\mu_0)' + \mathbf{A} \setminus \boldsymbol{\zeta} + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{H} + \mathbf{HH}, \\
\mathbf{X_0} &= (\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota}_0) \;\; (\mathbf{A} \setminus \boldsymbol{\iota}_0))\,, \\
\mathbf{X_1} &= \mathbf{I}_{(T\times T)}, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= (\tau_0 \;\mu_0)', \\
\widetilde{\mathbf{y}} &= \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{A} \setminus \boldsymbol{\zeta}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}$$

•*Local Linear Trend-SSOE*

$$\begin{aligned}
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \kappa_\mu\boldsymbol{\varepsilon} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HHc} \\
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota}_0\tau_0 + \boldsymbol{\iota}_0\mu_0 + \left(\kappa_\mu\mathbf{I}_{(T\times T)} + \kappa_\tau\mathbf{H} + \mathbf{HH}\right)\boldsymbol{\varepsilon} \\
\mathbf{HH}(\mathbf{A}\setminus\mathbf{y}) &= \left(\mathbf{H}(\mathbf{A}\setminus\boldsymbol{\iota}_0) \quad (\mathbf{A}\setminus\boldsymbol{\iota}_0)\right)(\tau_0\ \mu_0)' + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.}\begin{cases}
\mathbf{A} &= \kappa_\mu\mathbf{I}_{(T\times T)} + \kappa_\tau\mathbf{H} + \mathbf{HH}, \\
\mathbf{X_0} &= \left(\mathbf{H}(\mathbf{A}\setminus\boldsymbol{\iota}_0) \ \ (\mathbf{A}\setminus\boldsymbol{\iota}_0)\right), \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= (\tau_0\ \mu_0)', \\
\widetilde{\mathbf{y}} &= \mathbf{HH}(\mathbf{A}\setminus\mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}$$

I(1)-UC Models

•*MNZ-MSOE*

$$\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \boldsymbol{\eta} + \mathbf{Hc} \\
\mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi\boldsymbol{\iota}_0\tau_0 + \mathbf{H}_\phi\boldsymbol{\eta} + \mathbf{H}\boldsymbol{\varepsilon} \\
\mathbf{H}^{-1}\mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi(\mathbf{H}\setminus\boldsymbol{\iota}_0)\tau_0 + \mathbf{H}_\phi(\mathbf{H}\setminus\boldsymbol{\eta}) + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_0} & = \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\iota}_0), \\ \mathbf{X_1} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_2} & = \mathbf{H}_\phi, \\ \mathbf{z_0} & = \tau_0, \\ \widetilde{\mathbf{y}} & = \mathbf{H}_\phi \mathbf{y}, \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{0}_{T\times 1}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{H} \setminus \boldsymbol{\eta}. \end{cases}$$

•*MNZ-MSOE(UR)*

Recall from Section 2 that when we do not restrict the correlation between τ_t and c_t we specify $\eta_t = \eta_t^* + \kappa_\tau \varepsilon_t$, such that $\eta_t^* \sim \mathcal{N}(0, \sigma_\eta^2)$ and $\mathbb{C}\text{ov}(\varepsilon_t, \eta_t^*) = 0$. Therefore, using matrix notation we have:

$$\begin{aligned} \mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \boldsymbol{\eta}^* + \kappa_\tau\boldsymbol{\varepsilon} + \mathbf{Hc} \\ \mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi\boldsymbol{\iota}_0\tau_0 + \mathbf{H}_\phi\boldsymbol{\eta}^* + (\kappa_\tau\mathbf{H}_\phi + \mathbf{H})\boldsymbol{\varepsilon} \\ \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}) &= \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\iota}_0)\tau_0 + \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\eta}^*) + \boldsymbol{\varepsilon} \\ \widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \kappa_\tau\mathbf{H}_\phi + \mathbf{H}, \\ \mathbf{X_0} & = \mathbf{H}_\phi(H \setminus \boldsymbol{\iota}_0), \\ \mathbf{X_1} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_2} & = \mathbf{H}_\phi, \\ \mathbf{z_0} & = \tau_0, \\ \widetilde{\mathbf{y}} & = \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}), \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{0}_{T\times 1}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{A}\setminus\boldsymbol{\eta}^*. \end{cases}$$

•*MNZ-SSOE*

$$\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \kappa_\tau\boldsymbol{\varepsilon} + \mathbf{Hc} \\
\mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi\boldsymbol{\iota}_0\tau_0 + \left(\kappa_\tau\mathbf{H}_\phi + \mathbf{H}\right)\boldsymbol{\varepsilon} \\
\mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}) &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota}_0)\tau_0 + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{H}_{\mathbf{CE}} + \mathbf{H}, \\
\mathbf{X_0} &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota}_0), \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= \tau_0, \\
\widetilde{\mathbf{y}} &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}$$

•*Local Level-MSOE*

$$\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \boldsymbol{\eta} + \mathbf{H}\boldsymbol{\varepsilon} \\
\mathbf{y} &= \mathbf{H} \setminus \boldsymbol{\iota}_0\tau_0 + \mathbf{H}\backslash\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases}
\mathbf{A} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_0} &= \mathbf{H} \setminus \boldsymbol{\iota}_0, \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{I}_{(T\times T)}, \\
\mathbf{z_0} &= \tau_0, \\
\widetilde{\mathbf{y}} &= \mathbf{y}, \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{H}\backslash\boldsymbol{\eta}.
\end{cases}$$

•*Local Level-SSOE*

$$
\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \kappa_\tau\boldsymbol{\varepsilon} + \mathbf{H}\boldsymbol{\varepsilon} \\
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \left(\kappa_\tau\mathbf{I}_{(T\times T)} + \mathbf{H}\right)\boldsymbol{\varepsilon} \\
\mathbf{H}(\mathbf{A} \setminus \mathbf{y}) &= \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota}_0)\tau_0 + \boldsymbol{\varepsilon} \\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}
$$

$$
\text{s.t.} \begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{I}_{(T\times T)} + \mathbf{H}, \\
\mathbf{X_0} &= \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota}_0), \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= \tau_0, \\
\widetilde{\mathbf{y}} &= \mathbf{H}(\mathbf{A} \setminus \mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}
$$

Chapter 3

Monetary Policy and Trend Inflation: A Flexible Bayesian Modeling Framework

3.1 Introduction

This chapter develops an econometric framework to study the contribution of monetary policy as a driver of U.S. trend inflation. We detail model components and study the assess the role of systematic monetary policy as a determinant of trend inflation during the Great Inflation and Great Moderation episodes.

Trend inflation measures are commonly perceived as indicators of aggregate price behaviour in the long run. Estimates of trend inflation can, thus, be instrumental in a range of decision-making situations such as pricing long-term nominal bonds, setting nominal wage contracts as well as assessing the effectiveness of forward guidance about monetary policy rates. In light of this importance, recent years have witnessed a growing number of studies proposing different methods to estimate trend inflation as in, e.g., Stock and Watson [2007, 2015], Chan [2013], Chan et al. [2013, 2015] and Garnier et al. [2015]. To a certain extent, the unobserved components (UC) framework adopted in these studies suggests a consensus that estimation of trend inflation—defined in accordance with the Beveridge-Nelson concept of a trend for integrated time series[1]—should account for changes in the conditional first and second moments of the inflation process (e.g., changes in the conditional mean, volatility

[1] In other words, trend inflation is modeled as an unobserved (driftless) random walk process, hence consistent with the long run conditional expectation of inflation. Such choice purely reflects that the Beveridge-Nelson decomposition provides a well-established concept of trend in macroeconometrics. See Beveridge and Nelson [1981] for details. There are, however, alternative metrics for trend inflation such as survey-based measures of inflation expectations. The interested reader is referred to Clark and Doh [2014] for a comprehensive review on the different approaches to estimating trend inflation and their corresponding inflation (pseudo) out-of-sample performance.

and serial correlation of inflation). In addition, as a general rule, the results in these papers as well as in vector autoregression-based studies (see, for example, Cogley and Sargent [2001, 2005] and Cogley et al. [2010]) indicate that changes in U.S. inflation during the postwar period predominantly reflect movements in the trend (or permanent), rather than the transitory, component of inflation.[2]

Our study is largely motivated by the last point above. More specifically, concluding inflation dynamics are dominated by movements in its trend, while important, however, does not provide much insight into what drives trend inflation itself.[3] Indeed, although considerable attention has been paid to the econometric procedures for trend inflation inference, the sources of changes in trend inflation behaviour remain less well understood. Developing an econometric framework to shed light on this issue is the primary goal of this chapter.

Certainly, a formal account of the factors behind the rise and decline (in the level and volatility) of U.S. trend inflation in early 1970s and 1980s, respectively, should consider changes in the conduct of monetary policy as a plausible explanation. This view accords well with the perception that stabilization of long run inflation plays a chief role in modern central banking. In addition, the role of monetary policy as a nominal anchor to trend inflation is embedded in the design of many modern theoretical macroeconomic models (see, e.g., Woodford [2007], Ireland [2007], Cogley and Sbordone [2008], Castelnuovo [2010], Coibion and Gorodnichenko [2011] and Ascari and Sbordone [2014]). For example, in the context of a New Keynesian model with an interest rate-based policy rule, Woodford [2007] states, "the trend inflation rate... is also determined within the system: it corresponds to the central bank's target rate, incorporated into the policy rule...".

Alternatively, one could argue, for example, that trend inflation and the central bank's inflation target do not always coincide. A compelling case for such claim emerges if these variables reflect different views central bankers and market participants may have about long run inflation (see, for example, Binder [2015] and Kumar et al. [2015]). In a similar vein, one could claim that monetary policy represents one among several possible explanations for changes in trend inflation behaviour. In fact, the same rationale behind the discussion on the sources of the great moderation in business cycles (see, e.g., Kim and Nelson [1999], Stock and Watson [2003] and Galí and Gambetti [2009]) can be applied to investigate drivers of trend infla-

[2]The predominance of trend over transitory movements is not exclusive to U.S. inflation. See, e.g., Garnier et al. [2015] for cross-country comparisons.

[3]As Morley [2015] points out, its is important to formally investigate drivers of trend inflation that go beyond ex-post interpretations based on a broad sense of timing. For example, one such type of interpretation would be to attribute the rise in the level and volatility of U.S. trend inflation in the 1970s to the collapse of the Bretton Woods system, while the desinflationary episode in the early 1980s would be due to Paul Volcker's chairmanship of the Federal Reserve.

tion. For example, changes in the structure of the economy following technology improvements may have been equally relevant to determine postwar trend inflation dynamics. Competing alternatives could also include the role of fiscal policy in accordance with the fiscal theory of the price level.[4] Therefore, one might well wonder the extent to which changes in trend inflation indeed reflect changes in the conduct of monetary policy. To address this issue, and given the incipient stage of empirical research on trend inflation drivers, in this chapter we propose an econometric framework to evaluate the contribution of monetary policy to movements in U.S. trend inflation from the late 1960s until the onset of the financial crisis in the late 2000s.[5]

To be clear, our goal in this chapter is to develop a parsimonious and yet flexible framework to address the following empirical question: can movements in trend inflation be associated with changes in the conduct of systematic monetary policy? To do so, we combine two popular modeling strategies: (1) estimation of trend inflation using an unobserved components model similar to, e.g., Stock and Watson [2007], Kang et al. [2009], Chan et al. [2013] and Garnier et al. [2015]; (2) estimation of monetary policy rules in the spirit of Taylor [1993], but allowing for drifting coefficients akin to Cogley and Sargent [2005], Primiceri [2005], Boivin [2006], Kim and Nelson [2006] and Coibion and Gorodnichenko [2011].

The strategy outlined above allows us to construct a bivariate state space model where interactions between trend inflation and the parameters summarizing the evolution of systematic monetary policy are formally accounted and tested for. To be precise, we introduce cross-correlation between the latent stochastic components that characterize changes in the conduct of monetary policy and trend inflation. This is achieved by specifying the random process that drives trend inflation as a linear combination of two innovations: the first one is an exogenous term which we interpret as an all-encompassing non-policy shock to trend inflation.[6] The second innovation is identified as a systematic monetary policy shock to trend inflation. As such, it is represented by the same stochastic process that drives policy rule parameters and,

[4]The reader is referred to Leeper and Leith [2016] for a comprehensive discussion of macroeconomic models which incorporate interactions between monetary and fiscal policy in the determination of inflation and other macroeconomic outcomes.

[5]Our choice of sample period is justified on two grounds: (1) during the writing of this chapter, real-time forecasts in the Greenbook data set, used to construct a forward-looking policy rule, were available for a period spanning 1966Q1 to 2010Q4; (2) by focusing on a sample period up to the financial crisis we avoid potential pitfalls related to summarizing monetary policy actions solely by means of an interest rate rule when short-term nominal rates are near the zero lower bound (ZLB). For example, there is evidence that unconventional monetary policy tools adopted at the ZLB such as large scale purchase of assets can have significant impacts on Treasury term premia (see Meyer and Bomfim [2010], Gagnon et al. [2010] and **?**), which in turn can influence output and inflation (Kiley [2014]). Extending our framework to account for unconventional monetary policy is left for future work.

[6]For brevity, throughout this chapter we interchangeably use the word (non-)policy to refer to (non-)monetary policy effects.

consequently, introduces interdependence between changes in monetary policy and trend inflation.

A distinctive feature of our approach is the econometric methodology adopted to identify policy shocks to trend inflation. In particular, we explore the mapping between our econometric model and a structural representation of the monetary policy rule to distinguish between different policy drivers of trend inflation. This strategy allows us to overidentify a baseline state space model in different ways for each candidate policy shock underpinning trend inflation. As a result, evaluation of each identified policy shock and model comparison can be treated equivalently. In other words, assessing the suitability of different identification strategies becomes a matter of evaluating different overidentifying restrictions.

Model comparison is conducted using static and dynamic methods. The former entails computing the log marginal likelihood and deviance information criterion (DIC) (Spiegelhalter et al. [2002]) for each model following a particular identification strategy of the policy shock to trend inflation. Dynamic model comparison is carried out with a regime-switching approach. In practice, regime-switching allows for dynamic selection between overidentifying restrictions. The rationale behind this is to account for a more nuanced relationship between monetary policy and trend inflation, since the nature of monetary policy's contribution to trend inflation might reflect episode-contingent features.

Finally, the methodology adopted in this chapter to identify monetary policy shocks entails rank reduction of the covariance matrix for the state innovations. More precisely, rank reduction stems from different state variables being driven by a single source of randomness (i.e., perfectly correlated states) or some states becoming time invariant (i.e. state innovations with null variance).[7] To accommodate such features, we develop an efficient Markov Chain Monte Carlo (MCMC) algorithm which builds upon precision-based techniques in Chan and Jeliazkov [2009] and is generalized to encompass MCMC estimation of state space models with both full and reduced rank covariance matrices.

Overall, three main results emerge from our empirical analysis: first, monetary policy influence on trend inflation dynamics increased during the Great Moderation relative to the Great Inflation period. Second, despite such increase, non-policy factors represented more than 50% of variation behind trend inflation during both episodes. Third, we provide some evidence that monetary policy contribution to sta-

[7] From a statistical standpoint, one can perceive perfectly correlated states as an efficient way to use the data to estimate a large number of state variables without loosing modeling flexibility. Intuitively, if we have, say, five $T \times 1$ state vectors (i.e, $5T$ state variables) and a $T \times 1$ vector of observations, $\mathbf{y}$, then perfectly correlation between state vectors means that $\mathbf{y}$ can be used to estimate T instead of $5T$ state variables. Time variation of parameters in the model specification is, however, preserved.

bilize trend inflation during the 1980s can be characterized by a weaker reaction to changes in the output gap accompanied by an increased emphasis on inflation gap behaviour and inflation target adjustments.

The rest of this chapter is organized as follows: Section 3.2 presents the econometric framework. Section 3.3 discusses the two strategies proposed in this chapter to identify systematic monetary policy shocks to trend inflation. Estimation and an innovations-based parameterization of our models for posterior simulation are discussed in Section 3.4. Analysis of our results is presented in Section 3.5. Section 3.6 concludes.

3.2 A Cross-Correlated States Model for Inflation and the Policy Rule

We begin with a general representation of the bivariate spate space model for inflation, π_t, and the short-term nominal interest rate, r_t, adopted in this chapter. The idea is to provide a baseline we can refer back to when discussing how each identification strategy of the monetary policy contribution to trend inflation—presented in Section 3.3—overidentifies such baseline. The expressions below describe the measurement equations for π_t and r_t:

$$\begin{cases} \pi_t = \tau_t + e_{\pi,t}, \\ r_t = \beta_{0,t} + \beta_{1,t}\mathbb{E}_t(\pi_{t+h}) + \beta_{2,t}\mathbb{E}_t(y^{gap}_{t+h}) + \beta_{3,t}\mathbb{E}_t(m_{t+h}) + \beta_{4,t}r_{t-1} + e_{r,t}, \\ \\ \begin{bmatrix} e_{\pi,t} \\ e_{r,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\pi,j} & \sigma_{\pi r,j} \\ \sigma_{\pi r,j} & \sigma^2_{r,j} \end{bmatrix} \right), \\ \\ \text{for } j = 1 \text{ if } t \in (1967Q1, 1984Q1) \text{ and } j = 2 \text{ if } t \in (1984Q2, 2007Q4). \end{cases} \tag{3.2.1}$$

Inflation is defined as the sum of its trend (or permanent), τ_t, and transitory, $e_{\pi,t}$, components. Despite its simplicity, empirical relevance of such specification relative to more sophisticated variants have been demonstrated in, for example, Clark and Doh [2014] in the context of inflation forecasting.[8] As a robustness check, in Section 3.5.1 we show that allowing for serial correlation in the transitory component of π_t did not prove essential for our results.

[8] Intuitively, since stationary dynamics taper over time, the conditional expectation of future inflation is dominated by trend dynamics at longer horizons. Therefore, forecast performance at longer horizons emerges as a natural metric for selecting different models of trend inflation.

The second equation in (3.2.1) is a reduced form representation of a forward-looking Taylor-type policy rule with drifting coefficients, $\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$, and $\beta_{4,t}$. The structure underlying $\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$, and $\beta_{4,t}$ is discussed in the next section. Following Orphanides [2001, 2004] and Boivin [2006], estimation of a forward-looking policy rule relies on h-step-ahead conditional forecasts for inflation, $\mathbb{E}_t(\pi_{t+h})$, and the output gap, $\mathbb{E}_t(y^{gap}_{t+h})$. These forecasts are produced in real-time by the staff economists at the Federal Reserve Board of Governors and published in the Greenbook. Using real-time data is useful to replicate the information set available to policy makers when they are called to decide over interest rate levels. Also, to address empirical evidence provided in Ireland [2004] on the role of money for policy decisions, we augment the policy rule to account for growth in the stock of money, $\mathbb{E}_t(m_{t+h})$.[9] A detailed description of the data is provided in Section 3.5.1.

To accommodate changes in the volatility of and covariance between the innovations, $e_{\pi,t}$ and $e_{r,t}$, we allow for a structural break in all second moment parameters in the covariance matrix in (3.2.1). Following Kim and Nelson [1999], McConnell and Perez-Quiros [2000] and Stock and Watson [2003] the break date is set to the first quarter of 1984. Such break date reflects the possibility of a Great Inflation and Great Moderation volatility regime in our sample.[10]

Next, we present the specifications for the state variables, τ_t, $\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ and $\beta_{4,t}$. In keeping with previous studies (Cogley and Sargent [2001, 2005], Primiceri [2005], Boivin [2006], Kim and Nelson [2006], Stock and Watson [2007], Cogley et al. [2010] and Garnier et al. [2015]), trend inflation and the reduced form drifting coefficients in the policy reaction function are modeled as random walk processes.

[9] See also Canova and Ferroni [2012] for a recent application of a Taylor rule which incorporates money growth in the context of a New Keynesian dynamic stochastic general equilibrium model.

[10] See Boivin [2006] and Piger and Rasche [2008] for other examples of monetary policy and inflation modeling studies which also exogenously set similar break dates for the variance of the innovations. As discussed in in Section 3.5.5, our main findings are robust to the choice of break date.

Formally, we have:

$$
\begin{cases}
\tau_t = \tau_{t-1} + \overbrace{v_{\tau^*,t} + \gamma_j v_{r,t}}^{v_{\tau,t}} \quad \text{for } v_{r,t} \in \{v_{r_0,t},\ v_{r_1,t}, \cdots, v_{r_4,t}\}, \\
\beta_{0,t} = \beta_{0,t-1} + v_{r_0,t}, \\
\quad \vdots \\
\beta_{4,t} = \beta_{4,t-1} + v_{r_4,t}, \\
\begin{bmatrix} v_{\tau^*,t} \\ v_{r_0,t} \\ \vdots \\ v_{r_4,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{v_{\tau^*},j} & 0 & \cdots & 0 \\ 0 & \sigma^2_{v_{r_0},j} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{v_{r_4},j} \end{bmatrix} \right), \\
\text{for } j = 1 \text{ if } t \in (1967Q1, 1984Q1) \text{ and } j = 2 \text{ if } t \in (1984Q2, 2007Q4).
\end{cases}
\tag{3.2.2}
$$

As mentioned earlier, an important aspect of our approach is to allow for cross-correlation between the latent states that summarize changes in trend inflation and changes in the conduct of monetary policy. From the the first expression above, it is possible to verify that cross-correlation between trend inflation and some policy rule coefficient is enabled by defining the innovation, $v_{\tau,t}$, as a linear combination of two mutually uncorrelated innovations, $v_{\tau^*,t}$ and $v_{r,t}$. Since the latter is a common driver to both τ_t and some $\beta_{i,t} \in \{\beta_{0,t},\ \beta_{1,t}, \cdots, \beta_{4,t}\}$ for $i = 0, 1, ..., 4$, then trend inflation and some coefficient in the policy rule do not evolve independently. While details on the identification of different types of monetary policy shocks affecting τ_t are discussed in Section 3.3, suffice it to note here that for each identification strategy pursued in this chapter, a particular $v_{r,t} \in \{v_{r_0,t},\ v_{r_1,t}, \cdots,\ v_{r_4,t}\}$ enters the state equation for τ_t. In other words, $v_{r,t}$ captures a particular systematic monetary policy shock to the central bank's reaction function which in turn affects trend inflation dynamics.

The loading parameter, γ_j for $j = 1$ and 2, is introduced to dampen or magnify the effect of monetary policy changes on τ_t. Therefore, we interpret such parameter as a measure of the degree of transmission of the policy shock. Moreover, we also allow for a break in γ_j at 1984Q1 to capture potential differences in the transmission of a policy shock to τ_t between the Great Inflation and Great Moderation periods.

To conclude the description of the model, note that, as for the innovations in the measurement equations, we allow for a break at 1984Q1 in the volatility of the state

innovations. In particular, allowing for breaks in the conditional variance of the policy reaction coefficients accords well with narratives as in Clarida et al. [2000], who suggest that the decline in inflation volatility during the Great Moderation reflects a reduction in random monetary policy mistakes. In addition, by allowing for a break in the volatility of $v_{\tau^*,t}$ —which we treat as a non-policy shock to trend inflation— we do not rule out interpretations of a less volatile trend inflation during the Great Moderation being due to changes in the volatility of non-monetary policy related factors.[11]

3.3 A New Approach to Identify Systematic Monetary Policy Effects on Trend Inflation

In what follows, we discuss a method to identify systematic monetary policy shocks to trend inflation. To do so, we explore the mapping between the baseline model in (3.2.1)-(3.2.2) and a structural representation of the monetary policy rule. This allows us to overidentify the former and treat the identification exercise as a model selection one.

Next, we point out that the permanent and transitory decomposition of π_t in (3.2.1) accords well with monetary policy affecting inflation through both its systematic (long run) and non-systematic (short run) channels.[12] In particular, such decomposition of π_t is used to establish the common innovation ($v_{r,t}$ in (3.2.2)) between trend inflation and some policy coefficient as a systematic monetary policy shock transmitted to trend inflation. As demonstrated shortly, we propose six different types of systematic policy shocks which are classified according to following structural representation of the monetary policy rule:

$$\tilde{r}_t = r^* + \mathbb{E}_t(\pi_{t+h}) + \omega_1(\mathbb{E}_t(\pi_{t+h}) - \pi^*) + \omega_2\mathbb{E}_t(y^{gap}_{t+h}) + \omega_3\mathbb{E}_t(m_{t+h}), \tag{3.3.1}$$

$$r_t = (1-\rho)\tilde{r}_t + \rho r_{t-1} + e_{r,t}. \tag{3.3.2}$$

The equations above describe a forward-looking policy rule along the lines examined by Clarida et al. [2000], Orphanides [2001, 2004] and Boivin and Giannoni [2006]. Equation (3.3.1) is a Taylor rule, augmented to include policy responses to expected money growth, $\mathbb{E}_t(m_{t+h})$. As noted in Orphanides [2004], $\tilde{r}_t$ can be inter-

[11] See, for example, Bernanke and Mihov [1998] and Sims and Zha [2006]. Indeed, such studies suggest that the reduction in volatility during Great Moderation might purely reflect "good luck". The latter, translated as a change in the nature and size of the shocks that hit the economy during this period.

[12] See, for example, Bernanke et al. [1997] and Levin [2014] for a detailed discussion on the distinction between systematic and non-systematic monetary policy.

preted as a deterministic rule to set a notional target for the current quarter policy rate. In words, Equation (3.3.1) states that the determination of $\tilde{r}_t$ depends on the central bank's estimate of the real interest rate, r^*, the h-step-ahead conditional forecasts for inflation, the output gap and money growth as well as considerations about how much expected inflation might deviate from the central bank's inflation target, π^*, i.e., $\mathbb{E}_t(\pi_{t+h}) - \pi^*$. The latter commonly referred to as the inflation gap.[13]

Equation (3.3.2) reflects inertial behaviour from the monetary authority to achieve the target policy rate, $\tilde{r}_t$. Assuming $\rho \in [0, 1)$, Clarida et al. [2000] argue that (3.3.2) is indicative of the central bank's tendency to smooth interest rate movements.[14] Similar to Orphanides [2004] and Boivin [2006], the error term, $e_{r,t}$, reflects non-systematic monetary policy adjustments to r_t which are assumed to be independent of the inflation, output and money growth outlook in (3.3.1). Moreover, since correlation between $e_{r,t}$ and $e_{\pi,t}$ is allowed, if desired, one can factorize the covariance matrix in (3.2.1) accordingly to identify $e_{r,t}$ as a non-systematic monetary policy shock to inflation (see, e.g., Cogley and Sargent [2005], Primiceri [2005] and Boivin and Giannoni [2006]).[15] In this chapter, however, our focus is on the identification of systematic monetary policy shocks.

Turning to the identification of how systematic policy shocks can affect trend inflation, note that the system in (3.3.1)-(3.3.2) provides five candidate parameters to summarize the conduct of systematic monetary policy. Specifically, the central bank's inflation target, π^*, the policy smoothing parameter, ρ, and the three parameters, ω_1, ω_2 and ω_3, which respectively describe the degree of policy activism towards expected inflation, output and money growth dynamics.[16] Substituting equation (3.3.1) back into (3.3.2) then reveals a non-linear mapping between structural – $\{\pi^*,\ \omega_1,\ \omega_2,\ \omega_3,\ \rho\}$ – and reduced form policy parameters – $\{\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4\}$ –

[13] In this chapter, we do not pursue econometric identification of the real interest rate, r^*. An interesting extension to our work would be to specify a time varying r^* in order to study the joint determination of trend inflation and the real rate. One alternative could be to combine the framework presented here with the state space approach in Laubach and Williams [2003]). This would entail introducing an additional measurement and state equation to represent the IS-curve and the real interest rate, respectively.

[14] See Rudebusch [2002] for a contrary view on associating ρ with the central banks' preference to smooth policy rate adjustments.

[15] In Section 3.5.5, as a robustness check, we allow for lags of the policy rate to enter the inflation measurement equation. This accords well with the idea that non-systematic monetary policy shocks affect inflation with at least one quarter of lag (see, e.g., Primiceri [2005]). Our results for the role of systematic monetary policy as a driver of trend inflation, however, are essentially unaffected by such change.

[16] While —depending on the degree of nominal rigidities— the policy rate can affect the real rate in the short run, it is not conventional to think of changes in the conduct of systematic monetary policy in terms of central banks pursuing different levels of the long run real rate. Therefore, we leave r^* out of the set of structural policy parameters which characterize changes in systematic monetary policy. One can think of r^* as being determined by technology, preferences and other non-monetary policy related institutional factors.

as illustrated below:

$$r_t = \beta_0 + \beta_1 \mathbb{E}_t(\pi_{t+h}) + \beta_2 \mathbb{E}_t(y_{t+h}^{gap}) + \beta_3 \mathbb{E}_t(m_{t+h}) + \beta_4 r_{t-1} + e_{r,t}, \tag{3.3.3}$$

such that:

$$\begin{cases} \beta_0 = (1-\rho)(r^* - \omega_1 \pi^*), \\ \beta_1 = (1-\rho)(1+\omega_1), \\ \beta_2 = (1-\rho)\omega_2, \\ \beta_3 = (1-\rho)\omega_3, \\ \beta_4 = \rho. \end{cases} \tag{3.3.4}$$

As is well known, a common strategy in the existing literature on changes in systematic monetary policy is to study such changes by taking a time varying parameter (TVP) approach to some reduced form variant of the policy rule in (3.3.3) (see, e.g, Cogley and Sargent [2005], Primiceri [2005], Boivin [2006], Kim and Nelson [2006] and Coibion and Gorodnichenko [2011]). In these cases, each reduced form TVP is assumed to evolve independently. That is, the innovations driving each of the β's are uncorrelated as in the baseline model in (3.2.1)-(3.2.2). The mapping in (3.3.4), however, suggests that, if treated as random variables (such as in the case of TVP models), the reduced form policy coefficients could be correlated. Note, for example, that the structural parameters ρ and ω_1 map to more than one β_i for $i = 0, 1, ..., 4$. This opens the possibility to interpret variation in reduced form parameters as a result of a common structural policy driver. Moreover, if parameter variation is confined to some subset of $\{\pi^*,\ \omega_1,\ \omega_2,\ \omega_3,\ \rho\}$, then some, but not necessarily all β's, are time varying. In this chapter, we investigate such modeling alternatives to propose a new approach to identify systematic policy shocks to trend inflation.

In short, the following question underpins our identification strategy: given (3.3.3)-(3.3.4), what is the reduced form TVP policy rule specification that emerges if trend inflation reflects a particular change in the conduct of systematic monetary policy? To answer this question, we use the expressions in (3.3.4) to uncover reduced form parameter instability which is consistent with variation in one of the parameters in $\{\pi^*,\ \omega_1,\ \omega_2,\ \omega_3,\ \rho\}$.

Recall from (3.2.2) that we define the innovation to trend inflation as $v_{\tau,t} = v_{\tau^*,t} + \gamma_j v_{r,t}$ for $v_{r,t} \in \{v_{r_0,t},\ v_{r_1,t}, \cdots,\ v_{r_4,t}\}$. For simplicity, we assume $v_{\tau^*,t}$ captures non-policy shocks to trend inflation in a all-encompassing manner. Our goal, then, is to identify several types of systematic policy shocks and *individually* assess whether or not one of these shocks stand out as a key policy driver of trend inflation. That

is, while more than one structural policy parameter may have varied over time, by focusing on a case-by-case (structural) parameter variation approach, we attempt to detect systematic policy changes that are informative about the predominant policy shock underlying trend inflation.[17]

Examples 1, 2 and 3 below illustrate the cases where one postulates that changes behind trend inflation – in addition to non-policy shocks – reflect adjustments in the central bank's inflation target (i.e. a policy shock to π^*) or changes in the central bank's reaction function in terms of stabilizing the inflation and output gap (i.e. a policy shock to ω_1 and ω_2, respectively). For notational clarity, hereafter, whenever we refer to a specific monetary policy shock to trend inflation we replace the subscript, r_i for $i \in \{0, 1, 2, 3, 4\}$, in $v_{r_i,t}$, by the appropriate structural parameter in $\{\pi^*, \omega_1, \omega_2, \omega_3, \rho\}$ corresponding to the identified shock.

[17] Alternatively, one could pursue modeling time variation of the structural parameters directly. However, given the non-linear way such parameters enter the structural policy rule, setting a sampling algorithm, and consequently identification (which is ultimately our goal), would become less straightforward. Such issue would become even more pronounced in the regime switching strategy adopted in Section 3.3.1.2. For these reasons, we leave extending the approach suggested in this chapter to more non-linear environments for future research.

- **Example (1): inflation target adjustment shock** $(v_{\pi^*,t})$

$$r_t = \overbrace{(1-\rho)(r^* - \omega_1 \pi_t^*)}^{\beta_{0,t}} + \overbrace{(1-\rho)(1+\omega_1)}^{\beta_1} \mathbb{E}_t(\pi_{t+h}) +$$

$$+ \overbrace{(1-\rho)\omega_2}^{\beta_2} \mathbb{E}_t(y_{t+h}^{gap}) + \overbrace{(1-\rho)\omega_3}^{\beta_3} \mathbb{E}_t(m_{t+h}) + \overbrace{\rho}^{\beta_4} r_{t-1} + e_{r,t},$$

$$\text{States:} \begin{cases} \tau_t = \tau_{t-1} + v_{\tau^*,t} + \gamma_j v_{\pi^*,t}, \\ \beta_{0,t} = \beta_{0,t-1} + v_{\pi^*,t} \\ \beta_{i,t} = \beta_i \quad \text{for } i = 1, ..., 4 \text{ and } t = 1, ..., T. \end{cases} \tag{3.3.5a}$$

- **Example (2): inflation gap stabilization shock** $(v_{\omega_1,t})$

$$r_t = \overbrace{(1-\rho)(r^* - \omega_{1,t} \pi^*)}^{\beta_{0,t}} + \overbrace{(1-\rho)(1+\omega_{1,t})}^{\beta_{1,t}} \mathbb{E}_t(\pi_{t+h}) +$$

$$+ \overbrace{(1-\rho)\omega_2}^{\beta_2} \mathbb{E}_t(y_{t+h}^{gap}) + \overbrace{(1-\rho)\omega_3}^{\beta_3} \mathbb{E}_t(m_{t+h}) + \overbrace{\rho}^{\beta_4} r_{t-1} + e_{r,t},$$

$$\text{States:} \begin{cases} \tau_t = \tau_{t-1} + v_{\tau^*,t} + \gamma_j v_{\omega_1,t}, \\ \beta_{0,t} = \beta_{0,t-1} - v_{\omega_1,t} \\ \beta_{1,t} = \beta_{1,t-1} + v_{\omega_1,t} \\ \beta_{i,t} = \beta_i \quad \text{for } i = 2, 3 \text{ and } 4 \text{ and } t = 1, ..., T. \end{cases} \tag{3.3.5b}$$

- **Example (3): output gap stabilization shock,** $(v_{\omega_2,t})$

$$r_t = \overbrace{(1-\rho)(r^* - \omega_1 \pi^*)}^{\beta_0} + \overbrace{(1-\rho)(1+\omega_1)}^{\beta_1} \mathbb{E}_t(\pi_{t+h}) +$$

$$+ \overbrace{(1-\rho)\omega_{2,t}}^{\beta_{2,t}} \mathbb{E}_t(y_{t+h}^{gap}) + \overbrace{(1-\rho)\omega_3}^{\beta_3} \mathbb{E}_t(m_{t+h}) + \overbrace{\rho}^{\beta_4} r_{t-1} + e_{r,t}$$

$$\text{States:} \begin{cases} \tau_t = \tau_{t-1} + v_{\tau^*,t} + \gamma_j v_{\omega_2,t}, \\ \beta_{2,t} = \beta_{2,t-1} + v_{\omega_2,t} \\ \beta_{i,t} = \beta_i \quad \text{for } i = 0,1,3 \text{ and } 4 \text{ and } t = 1,...,T. \end{cases} \tag{3.3.5c}$$

Note that in (3.3.5b), as mentioned above, our identification approach accommodates correlation between reduced form policy coefficients. That is, since the same source of randomness is (implicitly) driving $\omega_{1,t}$, then, from (3.3.4), $\beta_{0,t}$ and $\beta_{1,t}$ become perfectly correlated. Also, since ω_1 enters the expressions for β_0 and β_1 in (3.3.4) preceded by a positive and negative sign respectively, the direction of the correlation between such states is preserved in the reduced form model. An analogous procedure is taken when we postulate $\omega_{1,t}$, $\omega_{2,t}$ and $\omega_{3,t}$ are driven by the same policy shock and when changes in systematic monetary policy are identified through a drifting ρ_t. Overall, our identification strategy overidentifies the baseline model by: introducing perfect correlation and time variation to a specific set of reduced form coefficients in the policy equation. Table (3.1), summarizes the six systematic monetary policy shocks proposed in this chapter as well as their corresponding overidentifying restrictions to reduced form drifting coefficients.

Table 3.1: List of models and their corresponding: (implied) structural and reduced form TVPs in the policy rule, identified systematic policy shock to trend inflation and linear overidentifying restrictions

Identifier	TVPs in the structural PR	TVPs in the reduced form PR
M_{R1}	π_t^*	$\beta_{0,t}$
M_{R2}	$\omega_{1,t}$	$\beta_{0,t}$ and $\beta_{1,t}$
M_{R3}	$\omega_{2,t}$	$\beta_{2,t}$
M_{R4}	$\omega_{3,t}$	$\beta_{3,t}$
M_{R5}	$\omega_{1,t}$, $\omega_{2,t}$ and $\omega_{3,t}$	$\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$
M_{R6}	ρ_t	All β's
Identifier	**Policy shock to τ_t (name)**	**Overidentifying restrictions to state innovations in (3.2.2)**
M_{R1}	$v_{\pi^*,t}$ (inflation target adjustment shock)	$v_{\pi^*,t} = v_{r_0,t}$ and $v_{r_i,t} = 0$ for $i = 1,2,3,4$
M_{R2}	$v_{\omega_1,t}$ (inflation gap stabilization shock)	$v_{\omega_1,t} = -v_{r_0,t} = v_{r_1,t}$ and $v_{r_i,t} = 0$ for $i = 2,3,4$
M_{R3}	$v_{\omega_2,t}$ (output gap stabilization shock)	$v_{\omega_2,t} = v_{r_2,t}$ and $v_{r_i,t} = 0$ for $i = 0,1,3,4$
M_{R4}	$v_{\omega_3,t}$ (money growth stabilization shock)	$v_{\omega_3,t} = v_{r_3,t}$ and $v_{r_i,t} = 0$ for $i = 0,1,2,4$
M_{R5}	$v_{\omega_{\{1,2,3\}},t}$ (policy stance shock)	$v_{\omega_{\{1,2,3\}},t} = -v_{r_0,t} = v_{r_i,t}$ for $i = 1,2,3$ and $v_{r_4,t} = 0$
M_{R6}	$v_{\rho,t}$ (policy inertia shock)	$v_{\rho,t} = v_{r_4,t} = -v_{r_i,t}$ for $i = 0,1,2,3$

Note: (a) TVP and PR stand for time varying parameter and policy rule, respectively; (b) reduced form TVPs are driven by a single source of randomness, identified as a systematic monetary policy shock to the structural policy parameter(s). Differences in the number of reduced form TVPs reflect the mapping between the structural and reduced form policy parameters as illustrated in (3.3.3)-(3.3.4).

3.3.1 Identification Selection

In this section we present two approaches to gauge evidence corresponding to each of the six identification strategies for policy shocks reported in Table 3.1. The first approach is a simple model comparison exercise based on four measures of the fit – the log marginal likelihood, the deviance information criteria (DIC), the Bayesian information criteria (BIC) and the expected value of the log likelihood – applied to the six variants in Table 3.1

The second approach entails dynamic model comparison. Specifically, we introduce regime switching to enable regime-dependent identification of systematic monetary policy drivers behind trend inflation. Operationally, to facilitate regime identification, we use the three best-fit models under the static model comparison case to set the number of regimes. That is, the switching occurs between the three best-fit identification strategies in Table 3.1, ranked according to the DIC, BIC, marginal likelihood and expected log likelihood values. Sections 3.3.1.1 and 3.3.1.2 discuss such strategies in more detail.

In what follows, we provide a matrix representation of the baseline model in (3.2.1)-(3.2.2). This is important, since estimation (and therefore selection of different identification strategies), as discussed in Section 3.4, is conducted using precision-based techniques as in Chan and Jeliazkov [2009].[18] Such an approach requires casting the system in (3.2.1)-(3.2.2) into a matrix form whereby observables (π_t and r_t) and state variables ($\tau_t, \beta_{0,t}, ..., \beta_{4,t}$) are stacked over $t = 1, ..., T$. This contrasts with the more traditional state space estimation procedure of reparameterizing (3.2.1)-(3.2.2) into a canonical AR(1) form in order to conduct inference with a Kalman filter.[19] As McCausland et al. [2011] point out, precision-based methods are more efficient than Kalman filter-based algorithms for Bayesian inference of state space models. In particular, precision-based estimation relies on exploring the band structure of high-dimensional matrices to expedite computation. As such, it involves sparse matrix routines which are computationally faster than Kalman filter-based recursions.[20]

Now, consider the following state space representation obtained by stacking ob-

[18] A detailed description of the posterior sampling algorithms developed in this chapter is given in Appendix 3.A.1 and 3.A.2.

[19] The reader is referred to Harvey [1989] and Durbin and Koopman [2012] for a textbook treatment of the Kalman filter.

[20] The reader is referred to Chan and Jeliazkov [2009] and McCausland et al. [2011] for a detailed discussion on precision-based methods for state simulation.

servations and state variables in (3.2.1)-(3.2.2) over $t = 1, ..., T$:

$$
\begin{aligned}
\boldsymbol{\pi} &= \boldsymbol{\tau} + \mathbf{e}_{\pi}, \\
\mathbf{r} &= \mathbf{I}_T \boldsymbol{\beta}_0 + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{X}_3 \boldsymbol{\beta}_3 + \mathbf{X}_4 \boldsymbol{\beta}_4 + \mathbf{e}_r, \\
\mathbf{H}\boldsymbol{\tau} &= \tau_0 \boldsymbol{\iota}_{\mathbf{0}} + \mathbf{v}_{\tau^*} + \mathbf{H}_{\gamma} \mathbf{v}_r \quad \text{for } i \in \{\mathbf{v}_{r_0},\ \mathbf{v}_{r_1}, \cdots,\ \mathbf{v}_{r_4}\}, \\
\mathbf{H}\boldsymbol{\beta}_0 &= \beta_{0,0} \boldsymbol{\iota}_{\mathbf{0}} + \mathbf{v}_{r_0} \\
&\vdots \\
\mathbf{H}\boldsymbol{\beta}_4 &= \beta_{4,0} \boldsymbol{\iota}_{\mathbf{0}} + \mathbf{v}_{r_4},
\end{aligned}
$$

where:

$$
\mathbf{H} = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & \ddots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & -1 & 1 & 0 \\
0 & 0 & \cdots & \cdots & -1 & 1
\end{bmatrix}, \quad
\mathbf{H}_{\gamma} = \begin{bmatrix}
\gamma_1 & 0 & 0 & \cdots & 0 & 0 \\
0 & \ddots & 0 & \cdots & 0 & 0 \\
0 & 0 & \gamma_1 & 0 & \vdots & 0 \\
\vdots & \vdots & 0 & \gamma_2 & 0 & \vdots \\
0 & 0 & \cdots & \cdots & \ddots & 0 \\
0 & 0 & \cdots & \cdots & 0 & \gamma_2
\end{bmatrix},
$$

$$
\mathbf{X}_i = \begin{bmatrix}
x_{i,1} & 0 & 0 & \cdots & 0 & 0 \\
0 & x_{i,2} & 0 & \cdots & 0 & 0 \\
0 & 0 & \ddots & 0 & \vdots & 0 \\
\vdots & \vdots & 0 & \ddots & 0 & \vdots \\
0 & 0 & \cdots & \cdots & x_{i,T-1} & 0 \\
0 & 0 & \cdots & \cdots & 0 & x_{i,T}
\end{bmatrix} \quad \text{for } x_{i,t} \in \left\{ \mathbb{E}_t(\pi_{t+h}),\ \mathbb{E}_t(y^{gap}_{t+h}),\ \mathbb{E}_t(m_{t+h}),\ r_{t-1} \right\},
$$

$\boldsymbol{\iota}_{\mathbf{0}} = (1, \cdots, 0)'$ and τ_0 and $\{\beta_{0,0}, \beta_{1,0},\ \beta_{2,0},\ \beta_{3,0},\ \beta_{4,0}\}$ denote state initialization parameters for trend inflation and policy rule coefficients respectively. Clearly, all matrices above are sparse.

To obtain a precision sampling consistent state space specification, let m denote the number of (continuous) state variables.[21] Then, using the Kronecker product operator, $\otimes$, and stacking the vectors for the measurement and state equations in the

[21] In our application, $m = 6$, since the set of continuous state variables is given by $\{\tau_t,\ \beta_{0,t},\ \beta_{1,t},\ \beta_{2,t},\ \beta_{3,t},\ \beta_{4,t}\}$. The term continuous is introduced here to distinguish (conditionally) Gaussian states (i.e., trend inflation and policy coefficients) from the discrete state variable which enters the baseline version in (3.2.1)-(3.2.2) when allowing for regime switching between identification strategies, as presented in Section 3.3.1.2.

previous system leads to a more compact expression:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Lambda} + \mathbf{e} \qquad \mathbf{e} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{e}}), \tag{3.3.6}$$

$$\mathbf{L}\boldsymbol{\Lambda} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{v} \qquad \mathbf{v} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{v}}), \tag{3.3.7}$$

where: $\mathbf{Y} = \begin{bmatrix} \boldsymbol{\pi} \\ \mathbf{r} \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_T & \mathbf{X}_1 & \cdots \ \mathbf{X}_4 \end{bmatrix}$, $\mathbf{L} = \mathbf{I}_m \otimes \mathbf{H}$, $\mathbf{F} = \mathbf{I}_m \otimes \boldsymbol{\iota}_{\mathbf{0}}$,

$\boldsymbol{\Lambda} = [\boldsymbol{\tau},\ \boldsymbol{\beta}_0, \cdots, \boldsymbol{\beta}_4]'$ and $\boldsymbol{\alpha} = (\tau_0,\ \beta_{0,0}, \cdots, \beta_{0,4})'$.

The state space form above is simply the baseline model in (3.2.1)-(3.2.2) expressed in matrix notation and, therefore, does not account for the overidentifying restrictions discussed in Section 3.3. To incorporate them, below we introduce a sparse restriction matrix, $\mathbf{R}_\gamma$, which accommodates both the static and dynamic identification selection approaches discussed next.

3.3.1.1 Method 1: Static Identification Selection

Remember from Section 3.3 that identification of systematic monetary policy shocks to trend inflation overidentifies the policy rule in (3.2.1) in two ways: some reduced form drifting coefficients in the policy rule in (3.2.1) become linearly dependent (perfectly correlated) or time variation is assigned to a specific subset of such coefficients in (3.2.1). To accommodate these, a restriction matrix, $\mathbf{R}_\gamma$, is introduced to the matrix representation in (3.3.7). This leads to the following state space form:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Lambda} + \mathbf{e} \qquad \mathbf{e} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{e}}), \tag{3.3.8}$$

$$\mathbf{L}\boldsymbol{\Lambda} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{R}_\gamma\mathbf{v} \qquad \mathbf{v} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{v}}). \tag{3.3.9}$$

Clearly, defining $\mathbf{R}_\gamma$ as an identify matrix returns the baseline model in (3.3.6)-(3.3.7). To illustrate how one can create a matrix $\mathbf{R}_\gamma$ which contains the linear overidentifying restrictions in Table 3.1, below we present the exact exclusion restrictions that go into $\mathbf{R}_\gamma$ to generate models M_{R1}, M_{R2} and M_{R3}. Again, these correspond to the identification of a systematic policy shock to trend inflation as an inflation target adjustment, inflation gap stabilization and output gap stabilization shock, respectively. Appendix 3.A.3 provides the structure of $\mathbf{R}_\gamma$ for all six models in Table 3.1.

• M_{R1} : inflation target adjustment shock ($\mathbf{v}_{\pi^*}$)

$$\mathbf{R}_\gamma \mathbf{v} = \begin{bmatrix} \mathbf{I}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & -\mathbf{I}_T & \mathbf{0}_T & \cdots & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & \ddots & & \mathbf{0}_T & & \vdots \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix} \begin{bmatrix} \mathbf{v}_{\tau^*} \\ \mathbf{v}_{\pi^*} \\ \mathbf{v}_{r_1} \\ \mathbf{v}_{r_2} \\ \mathbf{v}_{r_3} \\ \mathbf{v}_{r_4} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{\tau^*} + \mathbf{H}_\gamma \mathbf{v}_{\pi^*} \\ \mathbf{v}_{\pi^*} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{3.3.10}$$

• M_{R2} : inflation gap stabilization shock ($\mathbf{v}_{\omega_1}$)

$$\mathbf{R}_\gamma \mathbf{v} = \begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & -\mathbf{I}_T & \cdots & & \mathbf{0}_T \\ \vdots & \ddots & \mathbf{I}_T & & & \vdots \\ & & & \mathbf{0}_T & & \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix} \begin{bmatrix} \mathbf{v}_{\tau^*} \\ \mathbf{v}_{r_0} \\ \mathbf{v}_{\omega_1} \\ \mathbf{v}_{r_2} \\ \mathbf{v}_{r_3} \\ \mathbf{v}_{r_4} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{\tau^*} + \mathbf{H}_\gamma \mathbf{v}_{\omega_1} \\ -\mathbf{v}_{\omega_1} \\ \mathbf{v}_{\omega_1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{3.3.11}$$

• M_{R3} : output gap stabilization shock ($\mathbf{v}_{\omega_2}$)

$$\mathbf{R}_\gamma \mathbf{v} = \begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & \cdots & & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & & \ddots & \mathbf{I}_T & & \vdots \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix} \begin{bmatrix} \mathbf{v}_{\tau^*} \\ \mathbf{v}_{r_0} \\ \mathbf{v}_{r_1} \\ \mathbf{v}_{\omega_2} \\ \mathbf{v}_{r_3} \\ \mathbf{v}_{r_4} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{\tau^*} + \mathbf{H}_\gamma \mathbf{v}_{\omega_2} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{v}_{\omega_2} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{3.3.12}$$

It is not hard to verify that the expressions on the right-hand side above are consistent with the state representations in (3.3.5a)-(3.3.5c). For example, plugging the right-hand side of (3.3.10) into (3.3.9) returns state equations which satisfy the overidentifying restrictions stipulated for an inflation target adjustment shock. That is, as specified in (3.3.5a), τ_t is driven by a linear combination of $v_{\tau^*,t}$ and $v_{\pi^*,t}$; the intercept ($\beta_{0,t}$) in the reduced form policy rule is driven by $v_{\pi^*,t}$, while the remaining states become time invariant.

Also, it is worth remembering that when allowing for a different policy shock to drive trend inflation, we are not simply changing shock labels, so to speak, but actually altering the reduced form specification in (3.3.8)-(3.3.9). Consequently, given that each identification strategy in Table 3.1 affects a canonical representation of our baseline model, one should expect that changing the identification scheme also changes

the evaluation of the likelihood function. Likelihood-based metrics, thus, emerge as natural candidates to compare the fit of each identification procedure proposed in this chapter.

For (identification selection) method 1, we calculate four likelihood-driven metrics for model selection: the log marginal likelihood, the deviance information criteria (DIC), the Bayesian information criteria (BIC) and the expected value of the log likelihood. While details on how these measures are estimated are given in Appendix 3.A.4, suffice it to note here that such set of measures encompasses both the formal Bayesian marginal likelihood-based procedure for model comparison as well as measures which are less sensitive to prior elicitation. Hence contributing to the robustness of our identification selection exercise.[22]

3.3.1.2 Method 2: Identification Selection with Regime Switching

Method 1 provides an identification selection procedure which allows the detection of monetary policy drivers of trend inflation in a static fashion. While useful, such approach is silent on the possibility of a more nuanced relationship between trend inflation and the conduct of systematic monetary policy. For example, in Table 3.1 we suggest six different types of systematic monetary policy shocks to trend inflation. A natural question then is whether or not the identified policy shock underlying trend inflation varies over time. In other words, the behaviour of trend inflation might reflect time-contingent features of systematic monetary policy.

To address this issue, in this section we introduce a regime switching strategy to study regime-dependent policy drivers of trend inflation. The idea is a simple one: to allow for different policy shocks to represent the policy driver of trend inflation during certain (non-predetermined) periods. To be clear, regimes in this chapter are defined according to the particular monetary policy shock in Table 3.1 that contributes to movements in trend inflation over a certain period. However, one should keep in mind that, for example, an inflation target adjustment shock regime, does not imply that such shock is the only driver of trend inflation during such a regime. Recall from (3.2.2), that we assume trend inflation is driven by a policy and non-policy innovation. Since regime switching is applied only to the former, for convenience, we classify regimes accordingly.

Moreover, as shown later in Section 3.5.3, M_{R1}, M_{R2} and M_{R3} are the best-fitting models according to the four measures of the fit applied in method 1. As a result, to

[22]See Berg et al. [2004] and Eisenstat and Strachan [2014] for examples on the use of the DIC as a model selection criterion. See Morley and Piger [2012] and Luo and Startz [2014] for a recent application of the BIC as a quasi-Bayesian device for model selection and Carlin and Louis [1997] for a discussion on the interpretation of the expected value of the log likelihood as an empirical Bayesian metric for model comparison.

avoid overfitting the composition of shocks to trend inflation with multiple stochastic processes, we use such models to set the number of regimes at three. Method 1 can, thus, be perceived as an auxiliary strategy to improve regime identification by helping reduce the number of regimes one needs to specify.

For clarity, in what follows we use the state representation in (3.3.5a)-(3.3.5c) and the structure of $\mathbf{R}_\gamma$ in (3.3.10)-(3.3.12) to establish our approach for (identification selection) method 2. Now let $k_t \in \{1,\ 2,\ 3\}$ denote a multinomial discrete state variable following a first-order Markov process with transition probabilities $p_{\ell_{old},\ell} = \mathbb{P}\mathrm{r}\left(k_t = \ell | k_{t-1} = \ell_{old}\right)$ for $\ell_{old}, \ell = 1, 2, 3$. In addition, let $\mathbb{1}_{(k_t=\ell)}$ denote an indicator variable such that:

$$\mathbb{1}_{(k_t=\ell)} = \begin{cases} 1 & \text{if } k_t = \ell \text{ for } \begin{cases} \ell = 1,2,3 \text{ and} \\ t = 1, \cdots, T, \end{cases} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, a model which dynamically selects between the first three identification strategies in Table 3.1, i.e., M_{Rk_t} for $k_t \in \{1,\ 2,\ 3\}$, exhibits the same representation of the measurement equation in (3.2.1) but with state equations now given by:

$$\text{States: } \begin{cases} \tau_t = \tau_{t-1} + v_{\tau^*,t} + \mathbb{1}_{(k_t=1)}\gamma_j^{\pi^*} v_{\pi^*,t} + \mathbb{1}_{(k_t=2)}\gamma_j^{\omega_1} v_{\omega_1,t} + \mathbb{1}_{(k_t=3)}\gamma_j^{\omega_2} v_{\omega_2,t}, \\ \beta_{0,t} = \beta_{0,t-1} + \mathbb{1}_{(k_t=1)} v_{\pi^*,t} - \mathbb{1}_{(k_t=2)} v_{\omega_1,t}, \\ \beta_{1,t} = \beta_{1,t-1} + \mathbb{1}_{(k_t=2)} v_{\omega_1,t}, \\ \beta_{2,t} = \beta_{2,t-1} + \mathbb{1}_{(k_t=3)} v_{\omega_2,t}, \\ \beta_{i,t} = \beta_i, \\ \text{for } j = 1 \text{ if } t < T_{break} \text{ and } j = 2 \text{ otherwise and for } i = 3, 4;\ t = 1, ..., T, \end{cases} \tag{3.3.13}$$

where $\gamma_j^{\pi^*}$, $\gamma_j^{\omega_1}$ and $\gamma_j^{\omega_2}$ are policy shock-specific loading parameters. Since realizations of k_t are mutually exclusive, it is easy to check that the system above nests all state representations in (3.3.5a)-(3.3.5c).

Next, we define an appropriate $\mathbf{R}_\gamma$ which combines the restriction matrices in (3.3.10)-(3.3.12). To do so, we specify $\mathbf{R}_\gamma$ as a linear combination of the matrix structures in (3.3.10)-(3.3.12) as follows:

$$\mathbf{R}_\gamma = \sum_{\ell=1}^{3} \left(\mathbf{I}_m \otimes \mathbf{K}_\ell\right) \mathcal{R}_\ell, \tag{3.3.14}$$

where:

$$\mathbf{K}_\ell = \begin{bmatrix} \mathbb{1}_{(k_1=\ell)} & 0 & \cdots & 0 & 0 \\ 0 & \mathbb{1}_{(k_2=\ell)} & & & 0 \\ \vdots & 0 & \cdots & & \ddots \\ \vdots & & & \mathbb{1}_{(k_{T-1}=\ell)} & 0 \\ 0 & 0 & \cdots & 0 & \mathbb{1}_{(k_T=\ell)} \end{bmatrix} \tag{3.3.15}$$

and $\mathcal{R}_\ell \in \mathcal{R}$ for $\ell = 1, 2, 3$, such that:

$$\mathcal{R} = \left\{ \underset{(\mathcal{R}_1)}{\begin{bmatrix} \mathbf{I}_T & \mathbf{H}_{\gamma^{\pi^*}} & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{I}_T & \mathbf{0}_T & \cdots & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & \ddots & & \mathbf{0}_T & & \vdots \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}}, \underset{(\mathcal{R}_2)}{\begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{H}_{\gamma^{\omega_1}} & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & -\mathbf{I}_T & \cdots & & \mathbf{0}_T \\ \vdots & \ddots & \mathbf{I}_T & & & \vdots \\ & & & \mathbf{0}_T & & \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}}, \underset{(\mathcal{R}_3)}{\begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{H}_{\gamma^{\omega_2}} & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & \cdots & & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & & \ddots & \mathbf{I}_T & & \vdots \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}} \right\}.$$

That is, $\mathcal{R}$ is the set containing the restriciton matrices used to achieve the identification stragies corresponding to models M_{R1}, M_{R2} and M_{R3}, as shown in Section 3.1.1, while $\mathbf{I}_m \otimes \mathbf{K}_\ell$ works as a selection matrix of $\mathcal{R}_1$, $\mathcal{R}_2$ and $\mathcal{R}_3$. Of course, since k_t is time varying, the identification selection exercise becomes a dynamic one. To see it more clearly, note that combining $\mathbf{R}_\gamma$, as defined in (3.3.14), with a vector of innovations $\mathbf{v} = [\mathbf{v}'_{\tau^*}\ \mathbf{v}'_{\pi^*}\ \mathbf{v}'_{\omega_1}\ \mathbf{v}'_{\omega_2}\ \mathbf{v}'_{r_3}\ \mathbf{v}'_{r_4}]'$, yields:

$$\mathbf{R}_\gamma \mathbf{v} = \begin{bmatrix} \mathbf{v}_{\tau^*} + \mathbf{K_1}\mathbf{H}_{\gamma^{\pi^*}}\mathbf{v}_{\pi^*} + \mathbf{K_2}\mathbf{H}_{\gamma^{\omega_1}}\mathbf{v}_{\omega_1} + \mathbf{K_3}\mathbf{H}_{\gamma^{\omega_2}}\mathbf{v}_{\omega_2} \\ \mathbf{K_1}\mathbf{v}_{\pi^*} - \mathbf{K_2}\mathbf{v}_{\omega_1} \\ \mathbf{K_2}\mathbf{v}_{\omega_1} \\ \mathbf{K_3}\mathbf{v}_{\omega_2} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Note that if one of the extreme cases, $k_t = 1$ or $k_t = 2$ or $k_t = 3$ for all $t = 1, ...T$ occur, then, given the definition of $\mathbf{K}_\ell$ in (3.3.15), one simply recovers the vectors on the right-hand side of (3.3.10), (3.3.11) and (3.3.12), respectively. Therefore, any deviation from such cases introduces a richer time series structure of the policy shocks underpinning trend inflation.

Lastly, as argued in the case of method 1, because different shock identification

strategies entail different restrictions to the reduced form model in (3.2.1)-(3.2.2), one would expect the evaluation of likelihood function of M_{Rk_t} to change (for each k_t) accordingly. Consequently, common identification concerns within regime switching models, such as label-switching (see, for example, Frühwirth-Schnatter [2001], Hamilton et al. [2007] and Geweke [2007]), are less likely to be a problem in this study.

3.4 Bayesian Estimation

In this section we discuss our approach to estimate all models under the two identification selection methods presented in Section 3.3.1. Bayesian techniques are applied to estimate the states, parameters and hyperparameters (i.e., parameters associated with prior distributions). The choice of a Bayesian approach can be justified on several grounds. First, recall that under identification selection method 2, we introduced a discrete latent state variable, k_t, to allow for regime switching across different identification strategies. Since regime switching can occur at anytime and multiple times, such mechanism is equivalent to endogenously modeling an unknown number of structural breaks. By resorting to simulation-based approaches which entail sampling the unknown breaks (i.e., discrete states) from their (conditional) posterior distribution, Bayesian techniques propose a useful alternative to well known intractability issues of likelihood maximization when the number of breaks increases.[23]

Second, as recently suggested in Kim and Kim [2013a], Bayesian methods appear to be more robust than maximum likelihood-based inference to deal with the 'pile-up problem' (see, e.g., Sargan and Bhargava [1983] and Stock and Watson [1998]) in the context of state space models with an autoregressive moving average representation (such as ours).[24] Indeed, such problem can occur when the conditional variances of some (or all) state variables are close to zero. This is the case if, for example, systematic changes in the conduct of monetary policy are represented by small movements in policy coefficients.[25] As a result, the support of the likelihood function would have

[23]For example, the likelihood function typically becomes multimodal when the number of breaks is only limited by the sample size. In turn, multimodality of the likelihood surface makes finding a global maximum via numerical optimization routines cumbersome. Bayesian methods, in contrast, do not hinge on modal values but on the full support of the likelihood. The reader is referred to Bai and Perron [1998] and Boldea and Hall [2013] for examples of non-Bayesian approaches to endogenously modeling structural breaks. The asymptotic test distribution developed in such studies, however, cannot be readily applied to the class state space models used in this chapter.

[24]To be precise, by taking first differences of the measurement equations in (3.2.1) and using Granger's lemma (see Granger and Newbold [1986], p. 28-30), one can show that the state space representation in (3.2.1)-(3.2.2) has a vector autoregressive moving average model with exogenous regressors (VARMAX). Multivariate state space models having a VARMAX representation is a well known feature. See, for example, Casals et al. [2012].

[25]Alternatively, one could address small variation in policy coefficients by introducing regime switch-

point mass at the (null variance) boundary of the parameter space. As shown in Kim and Kim [2013b]), while maximum likelihood estimates typically 'pile-up' at boundary of the parameter space, even when the true parameter value is not at or close to the boundary, Bayesian estimators, by marginalizing over nuisance parameters, are relatively free of such problem.[26]

Finally, conceptually, state estimation accords well with the Bayesian paradigm. For example, the Kalman filter, widely used for estimation of state space models, essentially relies on recursive construction of first and second moments from a conditional posterior distribution during the updating and prediction steps of the filtering algorithm.[27]

Next, we describe the priors and the innovations-based parameterization adopted for posterior simulation. An outline of the posterior algorithms is also presented. Derivation details of such algorithms can be found in appendices 3.A.1 and 3.A.2.

3.4.1 Priors

Prior distributions are chosen to facilitate computation and, whenever possible, reflect similarities between our framework and previous studies using Bayesian methods to estimate macroeconometric models with drifting coefficients (e.g., Sims and Zha [1998], Cogley and Sargent [2001, 2005], Primiceri [2005], Koop et al. [2009]). In particular, we select independent conjugate prior distributions that commonly appear in empirical work. Such priors lead to conditional posterior distributions which one can readily simulate from.

Equation (3.2.2) shows that a Gaussian prior is assumed for the state variables in $\mathbf{\Lambda}$ (policy coefficients and trend inflation). Similarly, a Gaussian prior is placed for the state initialization vector, $\boldsymbol{\alpha} = (\tau_0,\ \beta_{0,0}, \cdots, \beta_{0,4})'$. In particular, prior means for the entries in $\boldsymbol{\alpha}$ corresponding to $(\beta_{0,0}, \cdots, \beta_{0,4})'$ are calibrated consistently with the structure of the policy rule in (3.3.4). Specifically, we set prior means $\mathbb{E}(\omega_1) = \mathbb{E}(\omega_2) = 0.5$, $\mathbb{E}(\omega_3) = 0$, $\mathbb{E}(\pi^*) = \mathbb{E}(r^*) = 2$ and $\mathbb{E}(\rho) = 0.5$. The values for $\mathbb{E}(\omega_1)$ and $\mathbb{E}(\omega_2)$ follow from the seminal presentation of the policy rule in Taylor [1993]. Setting $\mathbb{E}(\omega_3) = 0$, reflects a priori belief in favor of a police rule where, as in Taylor [1993], monetary policy reaction to money growth is not included. The value for $\mathbb{E}(\pi^*)$ and $\mathbb{E}(r^*)$ denote an inflation target and real interest rate, on average, of 2% respectively. This is a sensible approximation for the U.S. (see Laubach and Williams

ing in the variance of the state innovations along the lines of Koop et al. [2009].

[26] Stock and Watson [1998] propose a median unbiased estimator based on a local-to-zero transformation of the state innovation variance to mitigate the 'pile-up' effect on maximum likelihood estimation.

[27] Such claim holds regardless of whether inference is conducted via one or two-sided filtering. Once again, the reader is referred to Harvey [1989] and Durbin and Koopman [2012] for a textbook discussion of the Kalman filter.

[2003], Ireland [2007] and Ascari and Sbordone [2014]). From (3.3.2), $\mathbb{E}(\rho = 0.5)$ implies that, a priori, and on average, the monetary authority places equal weights between a Taylor rule-based estimate for $\tilde{r}_t$ and the one-quarter lagged value of r_t when setting the current quarter policy rate. Trend inflation is initialized letting $\tau_0 \sim \mathcal{N}(\pi_1,\ 1)$, i.e., using the first observation of inflation in our sample to tune the prior mean. Prior variance is also set to 1 for all entries in $(\beta_{0,0}, \cdots, \beta_{0,4})'$, which provides reasonably uninformative priors.

A Gaussian prior is also used for the loading parameters. In particular, we set $\gamma_j \sim \mathcal{N}(0,\ 1)$ for $j = 1, 2$, which implies that prior hyperparameters are calibrated equivalently for the Great Inflation and Great Moderation periods. As discussed in Section 2, γ_j captures the transmission of a policy shock to trend inflation. Hence, prior parameterization for γ_j is broadly in line with Primiceri [2005], Sims and Zha [2006] and Koop et al. [2009], inasmuch as these studies find that the transmission mechanism of policy shocks to inflation remained effectively unchanged during the Great Inflation and Great Moderation.

The prior for the measurement equation covariance matrix, $\Omega_{e,j}$, follows the standard practice of using a inverse whishart prior. We set $\Omega_{e,1} \sim \mathcal{IW}(I_2,\ 10)$ and $\Omega_{e,2} \sim \mathcal{IW}(0.32I_2,\ 10)$. In a similar vein, variances in the state covariance matrix are modeled with an inverse gamma prior. That is, $\sigma^2_{v_{i,1}} \sim \mathcal{IG}(10,\ 2.25)$ and $\sigma^2_{v_{i,2}} \sim \mathcal{IG}(10,\ 0.09)$ for $i = 0, 1, ..., 4$. Calibration of hyperparameters for both measurement and state covariance matrices is consistent with empirical evidence on the higher and lower volatility during the Great Inflation and Great Moderation, respectively (Kim and Nelson [1999], McConnell and Perez-Quiros [2000], Stock and Watson [2003] and Primiceri [2005]).[28]

Priors for models under identification selection method 2 are the same as described above. The only differences, obviously, are for the priors associated with the discrete state variable, k_t. As discussed in Section 3.3.1.2, k_t is modeled as Markovian process following a three-point multinomial distribution parameterized by a transition matrix:

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix}. \tag{3.4.1}$$

Following Chib [1996], we apply a Dirichlet prior for each row, $\mathbf{p}_\ell$ for $\ell = 1, 2, 3$, in $\mathbf{P}$. That is, $\mathbf{p}_1 \sim \mathcal{D}(90,\ 1,\ 1)$, $\mathbf{p}_2 \sim \mathcal{D}(1,\ 90,\ 1)$ $\mathbf{p}_3 \sim \mathcal{D}(1,\ 1,\ 90)$. Calibrating prior parameters for the transition probabilities in this fashion, implies that transitions

[28] The reader is refereed to Koop and Korobilis [2010] for a textbook discussion on the merits of using particular classes of priors for modeling the covariance matrix in multivariate time series models.

from one regime to another are expected – a priori – to be not so frequent. This is consistent with the notion of a particular monetary policy strategy to rein in trend inflation exhibiting some degree of persistence. Moreover, having nearly absorbing regimes (which is the case here) is useful to help capturing regime changes which are meaningful. In other words, stronger evidence is required from the data in order to favor one of the three overidentified variants discussed in Section 3.3.1.2.

3.4.2 An Innovations-Based Parameterization

When using Bayesian methods, estimation entails generating draws for the parameters of interest from a joint posterior distribution. To illustrate, let $\boldsymbol{\theta} = \{\boldsymbol{\Sigma_e},\ \boldsymbol{\Sigma_v},\ \boldsymbol{\alpha},\ \gamma\}$ denote the set of parameter and hyperparameters associated with models under identification selection method 1, where $\gamma = (\gamma_1,\ \gamma_2)'$. Therefore, the joint posterior distribution based on the state space representation in (3.3.8) and (3.3.9) can be expressed as:

$$f(\boldsymbol{\Lambda}, \boldsymbol{\theta}|\mathbf{Y}) = f(\boldsymbol{\Lambda}|\mathbf{Y}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}). \tag{3.4.2}$$

It should then be clear that if one could directly draw $\boldsymbol{\theta}$ from $f(\boldsymbol{\theta}|\mathbf{Y})$, conditional on such draw, forward filtering backward smoothing (FFBS) algorithms as in Carter and Kohn [1994] and Durbin and Koopman [2012] could be employed to sample $\boldsymbol{\Lambda}$ from $f(\boldsymbol{\Lambda}|\mathbf{Y}, \boldsymbol{\theta})$, thereby generating a draw for $\boldsymbol{\Lambda}$ and $\boldsymbol{\theta}$ from $f(\boldsymbol{\theta}, \boldsymbol{\Lambda}|\mathbf{Y})$. Repeating this procedure a number of times would generate a sequence of draws for $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$, hence allowing one to compute sample moments of interest.

Similarly, for models under identification selection method 2, letting $\boldsymbol{\theta} = \{\boldsymbol{\Sigma_e},\ \boldsymbol{\Sigma_v},\ \boldsymbol{\alpha},\ \gamma,\ \mathbf{P}\}$ and $\mathbf{k} = (k_1, \cdots, k_T)'$ (i.e., the vector of Markovian discrete states), the following joint posterior distribution ensues:

$$f(\boldsymbol{\Lambda}, \mathbf{k}, \boldsymbol{\theta}|\mathbf{Y}) = f(\boldsymbol{\Lambda}|\mathbf{Y}, \mathbf{k}, \boldsymbol{\theta}) f(\mathbf{k}|\mathbf{Y}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}). \tag{3.4.3}$$

Again, if one could directly sample from $f(\boldsymbol{\theta}|\mathbf{Y})$ the same procedure described above could be applied to sample $\boldsymbol{\Lambda}$ and $\boldsymbol{\theta}$, except now an additional step is required to draw $\mathbf{k}$ from $f(\mathbf{k}|\mathbf{Y}, \boldsymbol{\theta})$. The latter could be achieved, for example, with the FFBS samplers of Gerlach et al. [2000] or Fiorentini et al. [2014] which are designed to sample discrete states by integrating out continuous states (i.e., $\boldsymbol{\Lambda}$ is not a conditional factor in the sampling distribution of $\mathbf{k}$).

As is well known, however, direct sampling from $f(\boldsymbol{\theta}|\mathbf{Y})$ is often a hard task as it entails integration over a high-dimensional state space to obtain a sampling

expression for $\boldsymbol{\theta}$.[29] A popular solution to this problem is to use an MCMC algorithm which requires sequentially sampling from conditional posterior distributions which are of closed form, and therefore, easy to sample from. In other words, constructing a Gibbs sampler to conduct estimation. In particular, a common sampling strategy for Gaussian or conditionally Gaussian state space models (such as ours) is to sample the model parameters conditional on the states and then sample the states conditional on the model parameters. For example, an MCMC algorithm for models under method 2 can be summarized as the following two-step sampler:

$$f(\boldsymbol{\Lambda}, \mathbf{k}|\mathbf{Y}, \boldsymbol{\theta}) = f(\boldsymbol{\Lambda}|\mathbf{Y}, \mathbf{k}, \boldsymbol{\theta}) f(\mathbf{k}|\mathbf{Y}, \boldsymbol{\theta}), \tag{3.4.4}$$

$$f(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Lambda}, \mathbf{k}). \tag{3.4.5}$$

Clearly, for models under method 1, the first step would only require sampling $\boldsymbol{\Lambda}$.

In this chapter we adopt a variant of the sampling strategy above. To be precise, we integrate out the state vector $\boldsymbol{\Lambda}$ from our MCMC algorithm and sample the innovations vector, $\widetilde{\mathbf{v}} = \mathbf{L}^{-1}\mathbf{v}$, rather than $\boldsymbol{\Lambda}$.[30] The rationale behind that is to circumvent simulating $\boldsymbol{\theta}$ from the conditional posterior, $f(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Lambda}, \mathbf{k})$, whose kernel has a reduced rank covariance matrix (hence non-invertible).[31] Importantly, for state simulation in (3.4.4), rank reduction of the state innovations covariance matrix does not pose a sampling problem. Kalman filter based algorithms can handle state space models with reduced rank covariance matrix (see, for example, Gerlach et al. [2000], Giordani and Kohn [2008] and Casals et al. [2015]). However, for parameter sampling the filtering recursions in FFBS algorithms do not apply and well-defined distribution needs to be derived for proper sampling. To illustrate the issue more clearly, from Bayes' rule we have:

$$f(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Lambda}, \mathbf{k}) \propto f(\mathbf{Y}|\boldsymbol{\Lambda}, \mathbf{k}, \boldsymbol{\theta}) f(\boldsymbol{\Lambda}|\mathbf{k}, \boldsymbol{\theta}) f(\mathbf{k}|\boldsymbol{\theta}) f(\boldsymbol{\theta}), \tag{3.4.6}$$

$$f(\boldsymbol{\theta}|\mathbf{Y}, \widetilde{\mathbf{v}}, \mathbf{k}) \propto f(\mathbf{Y}|\widetilde{\mathbf{v}}, \mathbf{k}, \boldsymbol{\theta}) f(\widetilde{\mathbf{v}}|\mathbf{k}, \boldsymbol{\theta}) f(\mathbf{k}|\boldsymbol{\theta}) f(\boldsymbol{\theta}). \tag{3.4.7}$$

Comparing the two expressions above, it is easy to check that the second density on the right-hand side exhibits a reduced rank covariance matrix in (3.4.6) and a full rank

[29] The reader is referred to Koop et al. [2007], chapter 11, for a detailed presentation of Bayesian computation.

[30] Clearly, once we have a draw for $\widetilde{\mathbf{v}}$, then $\mathbf{v}$ can be recovered by setting $\mathbf{v} = \mathbf{L}\widetilde{\mathbf{v}}$. Then, plugging $\mathbf{v}$ into (3.3.9) returns the vector of states $\boldsymbol{\Lambda}$. Also, the change of variable, $\widetilde{\mathbf{v}} = \mathbf{L}^{-1}\mathbf{v}$, is introduced for simulation efficiency purposes. Sampling $\widetilde{\mathbf{v}}$ rather than $\mathbf{v}$ leads to a conditional posterior for the former which contains a high-dimensional sparse precision matrix. Hence, allowing to explore the gains in simulation speed from adopting precision sampling techniques. See Appendix 3.A.2 and Chan and Jeliazkov [2009] for more details.

[31] The term kernel is used here to describe the expression of the conditional posterior distribution which excludes the normalizing constant.

one in (3.4.7). In particular, from (3.3.9), we have $\boldsymbol{\Lambda}|\mathbf{k},\boldsymbol{\theta} \sim \mathcal{N}\left(\widetilde{\boldsymbol{\alpha}},\ \mathbf{L}^{-1}\mathbf{R}_\gamma\boldsymbol{\Sigma}_\mathbf{v}\mathbf{R}'_\gamma\mathbf{L}'^{-1}\right)$, where $\widetilde{\boldsymbol{\alpha}} = \mathbf{L}^{-1}\mathbf{F}\boldsymbol{\alpha}$. Rank reduction of the covariance matrix, $\mathbf{L}^{-1}\mathbf{R}_\gamma\boldsymbol{\Sigma}_\mathbf{v}\mathbf{R}'_\gamma\mathbf{L}'^{-1}$, thus, follows from the exclusion restrictions in $\mathbf{R}_\gamma$ which – conditional on a specific identification strategy of the policy shock to trend inflation – imply that some states in $\boldsymbol{\Lambda}$ are perfectly correlated while others have null conditional variance. In contrast, $\widetilde{\mathbf{v}}|\mathbf{k},\boldsymbol{\theta} \sim \mathcal{N}\left(\mathbf{0},\ \mathbf{L}^{-1}\boldsymbol{\Sigma}_\mathbf{v}\mathbf{L}'^{-1}\right)$ removes $\mathbf{R}_\gamma$ from the prior covariance matrix, which now has full rank covariance matrix, since $\boldsymbol{\Sigma}_\mathbf{v}$ is diagonal and $\mathbf{L}$ is a lower triangular Toeplitz matrix (see Pollock et al. [1999]). Furthermore, the remaining distributions above, i.e., the Gaussian likelihood, $f(\mathbf{Y}|\widetilde{\mathbf{v}},\mathbf{k},\boldsymbol{\theta})$, and priors defined in Section 3.4.1 for $f(\mathbf{k}|\boldsymbol{\theta})$ and $f(\boldsymbol{\theta})$ are well-defined. Hence, enabling estimation using MCMC sampling techniques.

Lastly, to sample $\boldsymbol{\theta}$ from (3.4.7) instead of (3.4.6), we need to recast the state space representation in (3.3.8) and (3.3.9) such that the new state vector is now given by $\widetilde{\mathbf{v}} = \mathbf{L}^{-1}\mathbf{v}$. Recall $\mathbf{L}^{-1}$ is block-diagonal and that $\mathbf{R}_\gamma$ is a square matrix whose submatrices are either null or diagonal matrices themselves. Thus, from (3.3.9) we have: $\boldsymbol{\Lambda} = \mathbf{L}^{-1}\mathbf{F}\boldsymbol{\alpha} + \mathbf{L}^{-1}\mathbf{R}_\gamma\mathbf{v} = \mathbf{L}^{-1}\mathbf{F}\boldsymbol{\alpha} + \mathbf{R}_\gamma\mathbf{L}^{-1}\mathbf{v} = \widetilde{\boldsymbol{\alpha}} + \mathbf{R}_\gamma\widetilde{\mathbf{v}}$. Substituting the latter into (3.3.8) and setting $\widetilde{\mathbf{Z}} = \mathbf{Z}\mathbf{R}_\mathbf{fl}$ yields:

$$\mathbf{Y} = \mathbf{Z}\widetilde{\boldsymbol{\alpha}} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{v}} + \mathbf{e}, \tag{3.4.8}$$

$$\widetilde{\mathbf{v}} \sim \mathcal{N}\left(\mathbf{0},\ \mathbf{L}^{-1}\boldsymbol{\Sigma}_\mathbf{v}\mathbf{L}'^{-1}\right). \tag{3.4.9}$$

Therefore, by transforming $\mathbf{R}_\gamma$ from a prior covariance matrix component to a conditional mean term in the measurement equation, the innovations-based parametrization proposed above treats $\mathbf{R}_\gamma$ similar to a linear restriction matrix in Wald-type test environments. However, here, $\mathbf{R}_\gamma$ selects the shocks that enter the likelihood function (i.e., appear in the measurement equations) or show up only as prior (i.e., the state equations). To be clear, if a vector of innovations is a key driver of trend inflation, then it should be part the likelihood function, thus, updating one's prior beliefs about trend inflation shocks. The innovations-based parametrization in (3.4.8)-(3.4.9) makes the latter point more explicit.

3.4.2.1 Posterior Simulation

Next, we briefly describe the set of conditional posterior distributions in the MCMC algorithms associated with identification selection methods 1 and 2 discussed in Section 3.3.1. For the sake of brevity, the exact expressions for the posterior moments for the sampling distributions below as well as technical details on the derivation of

the MCMC algorithms are postponed to appendices 3.A.1 and 3.A.2.

•**Algorithm 1 (Method 1)**:

Now let $\boldsymbol{\theta}_{-\mathbf{j}}$ denote elements in $\boldsymbol{\theta}$ other than $\mathbf{j}$. Also, let $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{pmatrix}$. Then, posterior draws for the models in Table 1 can be obtained by sequentially sampling from:

1. $f(\widetilde{\mathbf{v}}|\mathbf{Y}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\widetilde{\mathbf{v}}},\ \bar{\mathbf{D}}_{\widetilde{\mathbf{v}}}\right)$,

2. $f(\boldsymbol{\mu}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\boldsymbol{\mu}}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\boldsymbol{\mu}},\ \bar{\mathbf{D}}_{\boldsymbol{\mu}}\right)$,

3. $f(\Omega_{e,j}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\Omega_{e,j}}) \sim \mathcal{IW}\left(\bar{v}_{\Omega_{e,j}}\ \bar{\mathbf{W}}_{\Omega_{e,j}}\right)$,

4. $f(\sigma^2_{\tilde{v}_{i,j}}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\sigma^2_{\tilde{v}_{i,j}}}) \sim \mathcal{IG}\left(\bar{v}_{\sigma^2_{\tilde{v}_{i,j}}}\ \bar{S}_{\sigma^2_{\tilde{v}_{i,j}}}\right)$, for $i = \tau^*, r_0..., r_4$ and $j = 1, 2$.

•**Algorithm 2 (Method 2)**:

For method 2, our MCMC algorithm entails sampling from the four conditional posteriors above, but needs to be augmented to sample the vector of discrete states $\mathbf{k} = (k_1,\ k_2, \cdots, k_T)'$ and the transition probabilities in (3.4.1). To sample $\mathbf{k}$ we use the (two-sided) filtering algorithm of Chib [1996]. The rows $\mathbf{p}_\ell$ for $\ell = 1, 2, 3$ in the transition matrix are sampled from a Dirichlet conditional posterior. Formally, we have:

5. $f(\mathbf{k}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$,

6. $f(\mathbf{p}_\ell|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\mathbf{p}_\ell}) \sim \mathcal{D}\left(\bar{\alpha}_{\ell,1},\ \bar{\alpha}_{\ell,2},\ \bar{\alpha}_{\ell,3}\right)$.

3.5 Evaluation

In this section we empirically evaluate the identification selection methods proposed in Section 3.3.1. Models are estimated using a Gibbs sampler to draw from the set of conditional posteriors in Section 4.2.1. We run the Gibbs sampler 250000 times for each model and discard the first 25000 draws during the burn-in phase. The focus in this section is on the assessment of identified policy drivers of trend inflation. Appendix 3.A.5 reports additional results for posterior estimates of model parameters. Diagnostic metrics to evaluate the mixing properties of the MCMC algorithms

proposed in this chapter are also reported in Appendix 3.A.7.

3.5.1 Data

All models for identification selection methods 1 and 2 in Section 3.3.1 are estimated using quarterly data from 1967Q1 to 2007Q4. As mentioned before, to capture the real-time nature of monetary policy decisions, akin to Orphanides [2004] and Boivin [2006], we use Greenbook forecasts to construct measures of expected inflation ($\mathbb{E}(\pi_{t+h}|\Phi_t)$) and the output gap ($\mathbb{E}(y^{gap}_{t+h}|\Phi_t)$).

Expected inflation is given by the Greenbook forecast of the annualized percentage change in the GDP deflator between t and $t+h$.[32] Real-time output gap forecasts are not readily available from the Greenbook. Therefore, the expected output gap is calculated as the difference between the natural rate of unemployment and the forecasted unemployment rate. Defining the output gap in this fashion accords well with the conventional interpretation of positive and negative output gap values corresponding to an expansion and contraction in economic activity, respectively. Unemployment forecasts are extracted from the Greenbook, while the natural rate of unemployment is constructed as a historical moving average using all observations up to time t of the quarterly civilian unemployment rate. The latter is obtained from the Federal Reserve Bank of St. Louis website. We stress that the approach adopted here to construct an output gap metric follows Orphanides [2004] and Boivin [2006]. The reader is referred to such studies for justifications on the suitability of measuring the output gap in this fashion for a historical account of monetary policy. Also, forecasts in the Greenbook are not consistently available throughout the period. Therefore, whenever possible, we set the forecasting horizon (h) at three. When such observation is not available, we use the value from the nearest preceding horizon. Greenbook forecasts are based on the forecast made in the last month of each quarter.

Money growth forecasts are not available in the Greenbook. Therefore, to keep the real-time nature of the controls in the policy rule, we construct $\mathbb{E}(m_{t+h}|\Phi_t)$ using the first difference in logs of money supply (M2) obtained from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists. In particular, each observation is obtained from its contemporaneous vintage. This leads to a monthly time series. The quarterly counterpart is then calculated by averaging monthly realizations within the quarter.

Lastly, the dependent variables, i.e., the policy rate (r_t) and inflation (π_t) are

[32] Prior to October 1991 inflation forecasts are based on the GNP, instead of the GDP, deflator. This change, however, does not affect our results since it does not entail artificial shifts in inflation forecasts which could have been purely due to a change in the deflator metrics.

annualized quarterly observations. For inflation, given the quarterly GDP implicit price deflator figure, z_t, we set $\pi_t = 400 \log(z_t/z_{t-1})$ in (3.2.1). The policy rate is the federal funds rate directly obtained from Federal Reserve Bank of St. Louis website.

3.5.2 Bayesian Model Comparison Between Models with and without Cross-Correlated States

Before reporting results for the identification schemes proposed in Section 3.3, it is useful to provide some statistical evidence to support our state space models with cross-correlated policy coefficients and trend inflation. Therefore, in what follows we perform a formal Bayesian model comparison exercise to compare each model in Table 3.1 against its counterpart which assumes policy coefficients and trend inflation are uncorrelated. Given the decomposition of the reduced form innovation to trend inflation in (3.2.2), $v_{\tau,t} = v_{\tau^*,t} + \gamma_j v_{r,t}$ for $v_{r,t} \in \{v_{r_0,t},\ v_{r_1,t}, \cdots,\ v_{r_4,t}\}$ and $j = 1, 2$, one obtains uncorrelated trend inflation and policy coefficients by setting $\gamma_1 = \gamma_2 = 0$. Now, let M_{R0} denote such case. Then, a comparison based on the posterior odds between any model in Table 3.1 versus M_{R0} is defined as:

$$\frac{f\left(M_{Rs}|\mathbf{Y}\right)}{f\left(M_{R0}|\mathbf{Y}\right)} = \frac{f\left(\mathbf{Y}|M_{Rs}\right)}{f\left(\mathbf{Y}|M_{R0}\right)} \times \frac{f\left(M_{Rs}\right)}{f\left(M_{R0}\right)} \quad \text{for } s = 1, ..., 6.$$

Typically, one assumes that all models are equally likely a priori. This means $\frac{f(M_{Rs})}{f(M_{R0})} = 1$ which simplifies the posterior odds to the ratio of marginal likelihoods, $\frac{f(\mathbf{Y}|M_{Rs})}{f(\mathbf{Y}|M_{R0})}$, also known as the Bayes factor (BF). In our exercise, the larger the Bayes factor (or, equivalently, the larger the posterior odds) is, the stronger is the evidence towards cross-correlation between drifting policy coefficients and trend inflation (i.e., our approach). Moreover, since M_{R0} is nested in M_{Rs} for $s = 1, ..., 6$, the Bayes factor in favor of models exhibiting cross-correlated states can be computed using the Savage-Dickey density ratio (see, Verdinelli and Wasserman [1995]):

$$\text{BF}_{M_{Rs,R0}} = \frac{f(\gamma_1 = \gamma_2 = 0)}{f(\gamma_1 = \gamma_2 = 0|\mathbf{Y})} \quad \text{for } s = 1, ..., 6.$$

In other words, we need to evaluate the joint prior and posterior densities for γ_1 and γ_2 at $\gamma_1 = \gamma_2 = 0$. Evaluating the prior is immediate. It simply involves evaluating a bivariate normal density with prior hyperparameters calibrated as in Section 3.4.1. The posterior in the denominator is not of standard form, but it can be estimated via

Monte Carlo integration,

$$f(\widehat{\gamma_1 = \gamma_2 = 0}|\mathbf{Y}) = \frac{1}{L}\sum_{l=1}^{L} f(\gamma_1 = \gamma_1 = 0|\mathbf{Y}, \widetilde{\mathbf{v}}^{(l)}, \boldsymbol{\theta}^{(l)}_{-\{\gamma_1,\gamma_2\}}),$$

by summing over L (post burn-in) posterior draws of $\widetilde{\mathbf{v}}^{(l)}$ and $\boldsymbol{\theta}^{(l)}_{-\{\gamma_1,\gamma_2\}}$ from the MCMC algorithm 1 in Section 3.4.2.1. Also, since the prior , $f(\gamma_1, \gamma_2)$, and the likelihood, $f(\mathbf{Y}|\widetilde{\mathbf{v}}, \boldsymbol{\theta})$, are both normal densities, then, applying Bayes' rule and standard results from linear regression (see, e.g., Koop [2003]) leads to a conditional posterior, $f(\gamma_1, \gamma_1|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\{\gamma_1,\gamma_2\}})$, which also exhibits a normal density. As a result, evaluation of the summation above at $\gamma_1 = \gamma_2 = 0$ becomes straightforward.

Table 3.2: Model comparison results for models with cross-correlated versus uncorrelated trend inflation and policy coefficients. Values for $2\log\left(\text{BF}_{M_{Rs,R0}}\right)$ for $s = 1, ..., 6$ greater than 2 and 6 indicate positive and strong evidence of cross-correlation between changes in monetary policy and trend inflation. See Kass and Raftery [1995] and Raftery [1995] for details on using twice the log of the Bayes factor as a model selection tool.

Model	$2\log\left(\text{BF}_{M_{Rs,R0}}\right)$
M_{R1}	6.92
M_{R2}	5.61
M_{R3}	6.12
M_{R4}	4.05
M_{R5}	3.52
M_{R6}	3.31

Table 3.2 reports model comparison results based on the $2\log\left(\text{BF}_{M_{Rs,R0}}\right)$ metric for $s = 1, ..., 6$ suggested in Kass and Raftery [1995] and Raftery [1995]. Our results show that the data favor cross-correlation between policy rule coefficients and trend inflation over a variant which assumes such states are orthogonal. Evidence of cross-correlation holds for all the identification strategies proposed in Section 3.3. Nonetheless, such evidence is stronger for the first three variants in Table 3.1. In the next sections we look in some more detail at these particular cases.

3.5.3 Results for Static Identification Selection

Having established evidence in favor of cross-correlation between policy coefficients and trend inflation, we now turn to the selection and evaluation of the identification strategies summarized in Table 3.1. Recall from Section 3.3 that each identification scheme of the policy shock to trend inflation corresponds to an overidentified variant of the system in (3.2.1)-(3.2.2). Thus, one can treat identification selection as a model

selection exercise. This is the strategy we pursue in this section.

Table 3.3 reports model fit results based on estimates of the DIC, the expected log likelihood, BIC and log marginal likelihood for each model in Table (3.1). Again, the reader is referred to Appendix 3.A.4 for a detailed description of how these metrics are calculated. Suffice it to note here that a smaller DIC indicates a better fit, while the converse applies for the other three measures.

Overall, results in Table 3.3 point to two main findings: (a) M_{R1} is our best-fit model under identification selection method 1. It outperforms all other variants according to three out of the four selection criteria adopted in this chapter. (b) There is a clear distinction in model fit between two sets of models, namely, $\{M_{R1},\ M_{R2},\ M_{R3}\}$ and $\{M_{R4},\ M_{R5},\ M_{R6}\}$. The former contains the three best-fit models according to any of the four metrics of model selection in Table 3.3. For example, the value for twice the log of the Bayes factor between the worst and best-fit models in the first and second sets respectively (i.e., $2\log\left(\text{BF}_{M_{R3,R4}}\right)$), is 90.46. This suggests very strong evidence in favor of M_{R3} against M_{R4}. Therefore, to facilitate regime identification, in the next section we focus on $\{M_{R1},\ M_{R2},\ M_{R3}\}$ in the empirical application of identification selection method 2, as discussed in Section 3.3.1.2.

Table 3.3: Deviance information criterion (DIC), expected log likelihood , BIC and log marginal likelihood for all models under under identification selection method 1

Model	DIC	$\mathbb{E}$(log likelihood)	BIC	log marginal likelihood
M_{R1}	**1087.8**	**-391.94**	**-303.98**	-250.94
M_{R2}	1139.3	-405.23	-307.01	**-236.96**
M_{R3}	1094.0	-395.75	-310.73	-253.54
M_{R4}	1298.1	-501.04	-419.23	-298.77
M_{R5}	1426.6	-559.86	-472.64	-309.07
M_{R6}	1594.8	-573.25	-489.30	-342.31

Note: All measures for model comparison are computed using a variant of the likelihood function with marginalization over state innovations. See appendix 3.A.4 for details. The log marginal likelihood is simulated using the modified harmonic mean approach with tail truncation. See Geweke [1999] for details. Results above are based on the sample period from 1967Q1 through 2007Q4 and allowing for a break in the variance of all innovations at 1984Q1. Results in bold denote the best-fit model according to each selection criterion.

Another important result that emerges from method 1 is presented in Table 3.4. It reports variance decomposition results for policy versus non-policy shocks contribution to trend inflation dynamics.[33] We find that more than 50% of the variation

[33]Variance decomposition results for a particular policy and a non-policy shock are respectively calculated as follows: $\frac{\mathbb{V}\text{ar}(\gamma_j v_{r_i,t})}{\mathbb{V}\text{ar}(v_{\tau^*,t}+\gamma_j v_{r_i,t})} = \frac{\gamma_j^2\sigma^2_{v_{r_i}}}{\sigma^2_{\tau^*}+\gamma_j^2\sigma^2_{v_{r_i}}}$ and $\frac{\mathbb{V}\text{ar}(v_{\tau^*,t})}{\mathbb{V}\text{ar}(v_{\tau^*,t}+\gamma_j v_{r_i,t})} = \frac{\sigma^2_{v_{\tau^*}}}{\sigma^2_{\tau^*}+\gamma_j^2\sigma^2_{v_{r_i}}}$ for $i = \{0,\ 1,\ 2,\ 3,\ 4\}$ and $j = 1,\ 2$. See Table 3.1 for details.

in trend inflation during both the Great Inflation and Great Moderation are due to exogenous non-policy shocks. In terms of the predominance of non-policy shocks driving inflation, our results are comparable with other studies such as Smets and Wouters [2007], Galí and Gambetti [2009] and Canova and Ferroni [2012]. Nonetheless, by assessing a range of policy shocks to a particular component of inflation (trend inflation) over two distinct episodes, our results reveal interesting aspects about the relationship between monetary policy and long run inflation. In particular, having M_{R1} as our best-fit model indicates that the adoption of an inflation targeting strategy – while not being the sole reason behind a well-behaved trend inflation during the Great Moderation – enhanced monetary policy ability to rein in long run inflation expectations.

Before turning to results from method 2 it is useful to assess if the particular sequence of inflation target adjustment shocks we estimated using the MCMC algorithm 1 returns a sensible measure of inflation target. To do that, we feed posterior draws of $\{v_{\pi^*,t}\}_{t=1}^{T}$ into an implied random walk representation of the inflation target given by $\pi_t^* = \pi_{t-1}^* + v_{\pi^*,t}$.[34] Figure 3.1 shows the resulting time series (posterior medians). Our implied measure of the inflation target captures the main features reported in other studies (see, for example, Ireland [2007] and Coibion and Gorodnichenko [2011]). In particular, it displays a gradual decrease pattern starting in the Volcker disinflation period with the target moving from 5% in the early 1980's to around 2% during the Great Moderation.

3.5.4 Results for Dynamic Identification Selection

As discussed before, method 1, while informative, is silent about possible shifts in the nature of policy shocks underlying trend inflation. To address this, we now report identification selection results based on the regime switching procedure discussed in Section 3.3.1.2, i.e., method 2. Recall such an approach introduces a multinomial discrete state variable, $k_t \in \{1,\ 2,\ 3\}$, which selects one of the three best-fit models under method 1 $(M_{R1},\ M_{R2},\ M_{R3})$ at each point $t = 1, ..., T$. Moreover, recall under method 2 each model corresponds to a particular regime described by the identified monetary policy shock to trend inflation.[35]

Figure 3.2 reports the posterior probabilities ($\mathbb{P}\mathrm{r}\left(k_t = i|\mathbf{Y}\right)$ for $i = 1, 2, 3$ and $t = 1, ..., T$) for an inflation target adjustment, inflation gap stabilization and output gap stabilization regime.[36] To a certain extent, our results in this section corroborate our

[34] We initialize the random walk representation of π_t^* with a diffuse prior given by $\pi_0^* \sim \mathcal{U}\left(2,\ 6\right)$.

[35] Recall regimes are informative about a specific systematic monetary policy shock underlying trend inflation in addition to non-policy shocks.

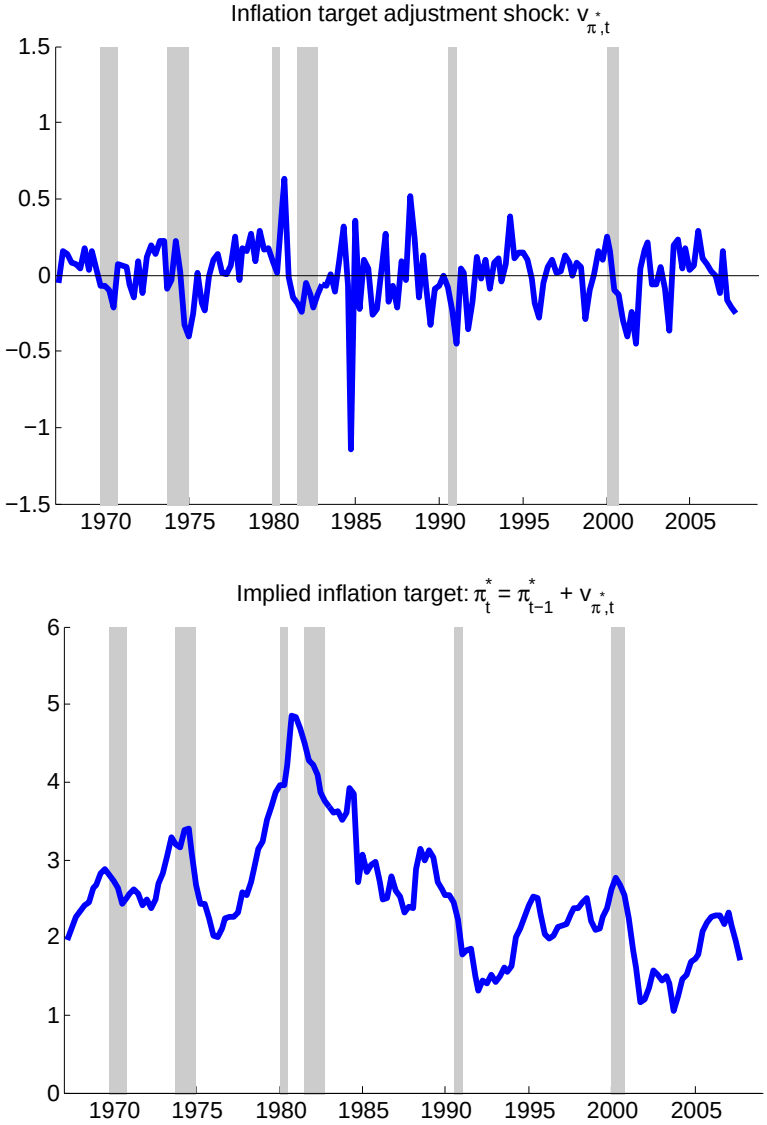
[36] Posterior regime probabilities are constructed by calculating the frequency each realization of $k_t \in$

Table 3.4: Variance decomposition results for policy and non-policy shocks shocks (posterior medians) to trend inflation for all models under identification selection method 1

Model	Policy Shock	
	Great Inflation	Great Moderation
M_{R1}	**31.74**%	**48.87**%
M_{R2}	8.33%	18.50%
M_{R3}	28.31%	49.81%
M_{R4}	10.22%	13.82%
M_{R5}	9.39%	9.13%
M_{R6}	9.19%	7.02%
Model	Non-Policy Shock	
	Great Inflation	Great Moderation
M_{R1}	**69.26**%	**51.23**%
M_{R2}	91.67%	81.50%
M_{R3}	71.69%	50.19%
M_{R4}	89.78%	86.18%
M_{R5}	90.61%	90.77%
M_{R6}	90.81%	92.98%

Note: Results based on the sample period from 1967Q1 through 2007Q4 and allowing for a break in the variance of measurement and state innovations at 1984Q1. Results in bold denote the best-fit model ranked according to the four selection criteria in Table 3.3.

Figure 3.1: Estimated sequence of systematic monetary policy shocks (posterior medians) and implied structural policy parameter for M_{R1}. The best-fit model under identification selection method 1. The structural policy parameter below is recovered by feeding the identified policy shock to the random walk representation of π_t^*. Shaded regions denote recession periods as recorded by the NBER.



findings from method 1. Inflation target adjustment shocks still stand out as the key policy driver of trend inflation. Specifically, the posterior probability associated with such shocks is around 70% and close to 100% at the start of the sample and throughout the Great Moderation, respectively. Nevertheless, method 2 suggests a more nuanced picture in terms of policy contributions to trend inflation during the transition from the Great Inflation to the Great Moderation. In particular, the period from 1975 to 1984 is marked by an alternation between inflation gap and output gap stabilization shocks as the main systematic monetary policy influences behind trend inflation.

To assess what such shocks entail for the dynamics of structural policy parameters, as in the last section, we feed the sequence of shocks, $\{v_{\pi^*,t},\ v_{\omega_1,t},\ v_{\omega_2,t}\}_{t=1}^{T}$ (see Figure 3.3), obtained from MCMC algorithm 2 into the implied random walk representation of each structural policy parameter, π_t^*, $\omega_{1,t}$ and $\omega_{2,t}$, accordingly. The resulting time series (posterior medians) are shown in Figure 3.4.[37]

Overall, our results are broadly aligned with previous studies such as Orphanides [2001], Boivin [2006], Kim and Nelson [2006] and Coibion and Gorodnichenko [2011]. In particular, we find that monetary policy reaction to expected output gap movements declined around the mid 1970s. This was accompanied by an upward trend in policy emphasis to react to expected deviations of inflation from its target. Furthermore, similar to Coibion and Gorodnichenko [2011] we find that the Taylor principle, captured here as $1 + \omega_t > 1$, is satisfied throughout the sample period. For the implied inflation target, we observe a steady decline from around 4% to 2.5% around 1987.

Taken together, these results indicate that systematic monetary policy contribution to bring down trend inflation between the late 1970s and early 1980s can be mapped to an increase in the degree of policy activism towards expected inflation relative to the output gap outlook. During the Great Moderation, when inflation expectations were less unhinged, monetary policy adjustments to its target inflation rate and trend inflation appeared to be better aligned.

Lastly, one should not expect the dynamics of the structural policy parameters reported in Figure 3.4 to mimic exactly the dynamics observed in other studies of TVP policy rules where all coefficients are driven by orthogonal innovations (e.g., Boivin [2006], Kim and Nelson [2006] and Coibion and Gorodnichenko [2011]). As discussed in Section 3.4.2, within our exercise if a particular shock is not a key driver of trend inflation then it does not enter the likelihood function (i.e., it is not up-

$\{1,\ 2,\ 3\}$ occurs given all MCMC draws from algorithm 2 in Section 3.4.2.1.

[37] As in Section 3.5.3, we initialize random walk representations of structural policy parameters with a diffuse prior. More precisely, we adopt the following initial conditions $\pi_0^* \sim \mathcal{U}(2,\ 6)$ and $\omega_{i,0} \sim \mathcal{U}(0,\ 2)$ for $i = 1, 2$.

dated). As a result, such shock is simply sampled from its prior density which has zero mean and median. This is reflected in Figure 3.4, which shows that parameters become virtually time invariant when their corresponding policy shock regime probabilities in Figure 3.2 are low. With that in mind, the fact that our results are qualitatively comparable with the ones obtained from studies adopting considerably different modeling frameworks is reassuring.

Figure 3.2: Posterior probabilities (posterior medians) for each type of systematic monetary policy shock driving trend inflation under method 2: regime switching between the three best-fit identification strategies under method 1. Shaded regions denote recession periods as recorded by the NBER.

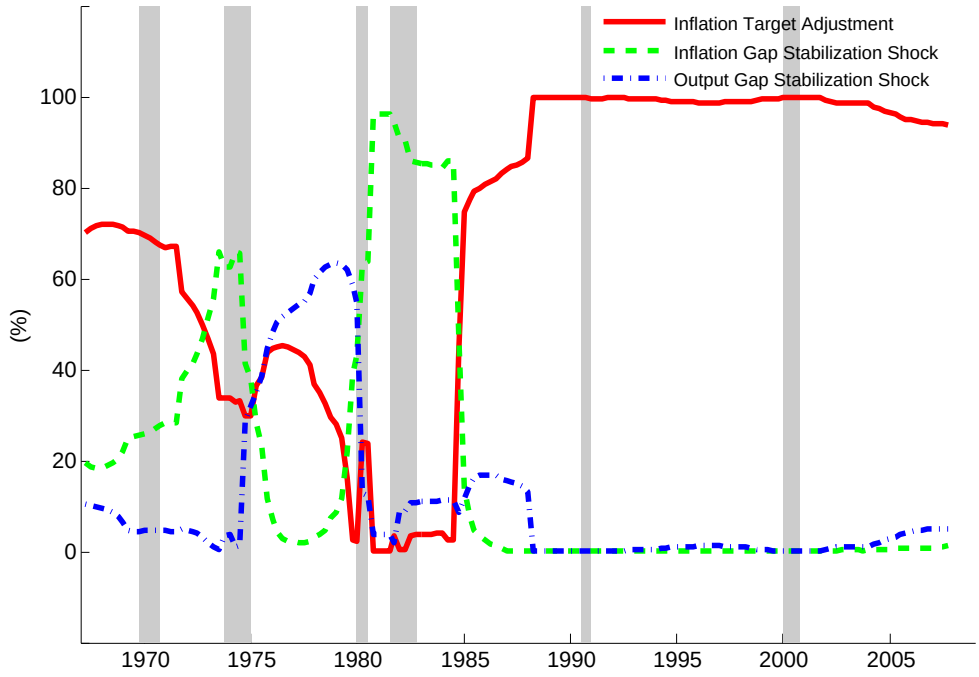


Figure 3.3: Estimated sequence of systematic monetary policy shocks (posterior medians) to trend inflation under identification selection method 2: regime switching between the three best-fit identification strategies under method 1. Shaded regions denote recession periods as recorded by the NBER.

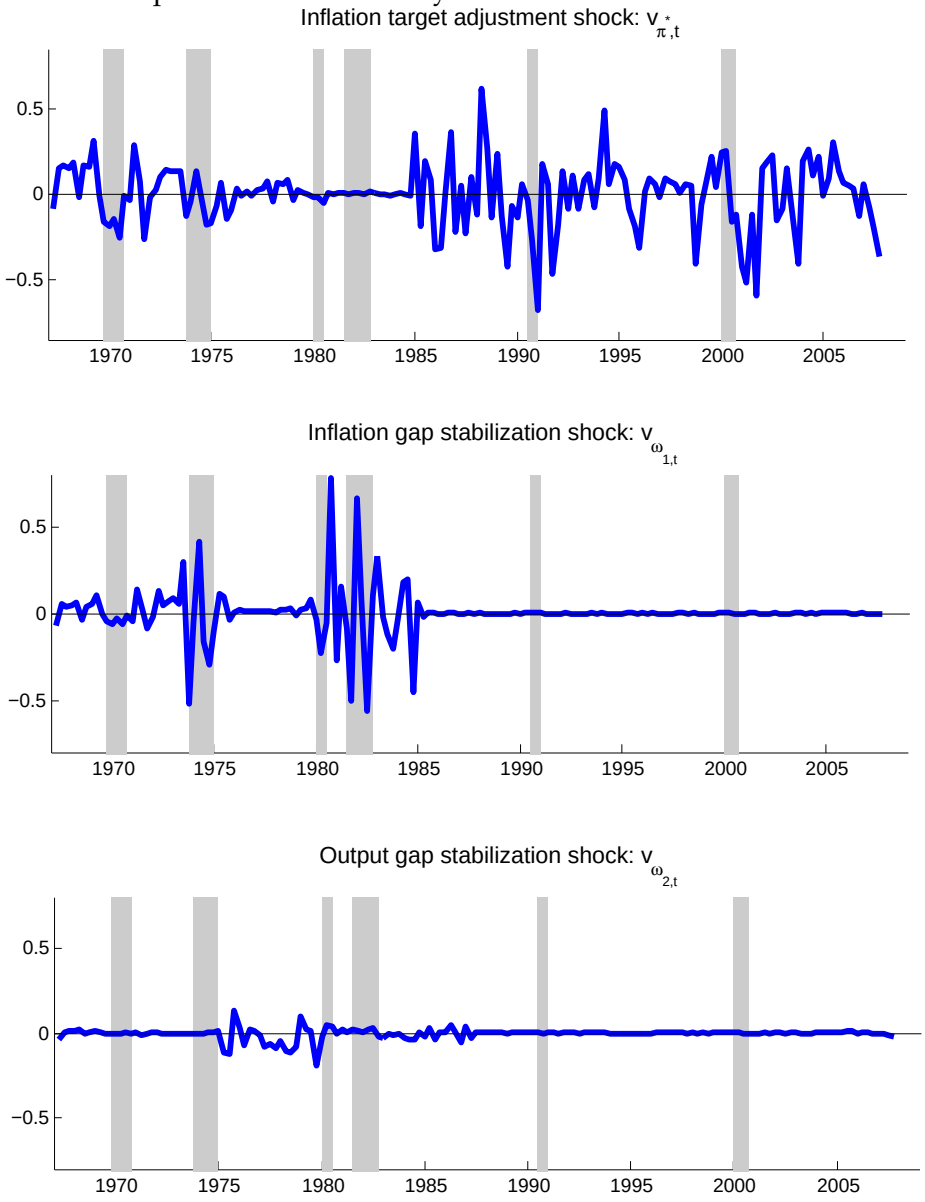
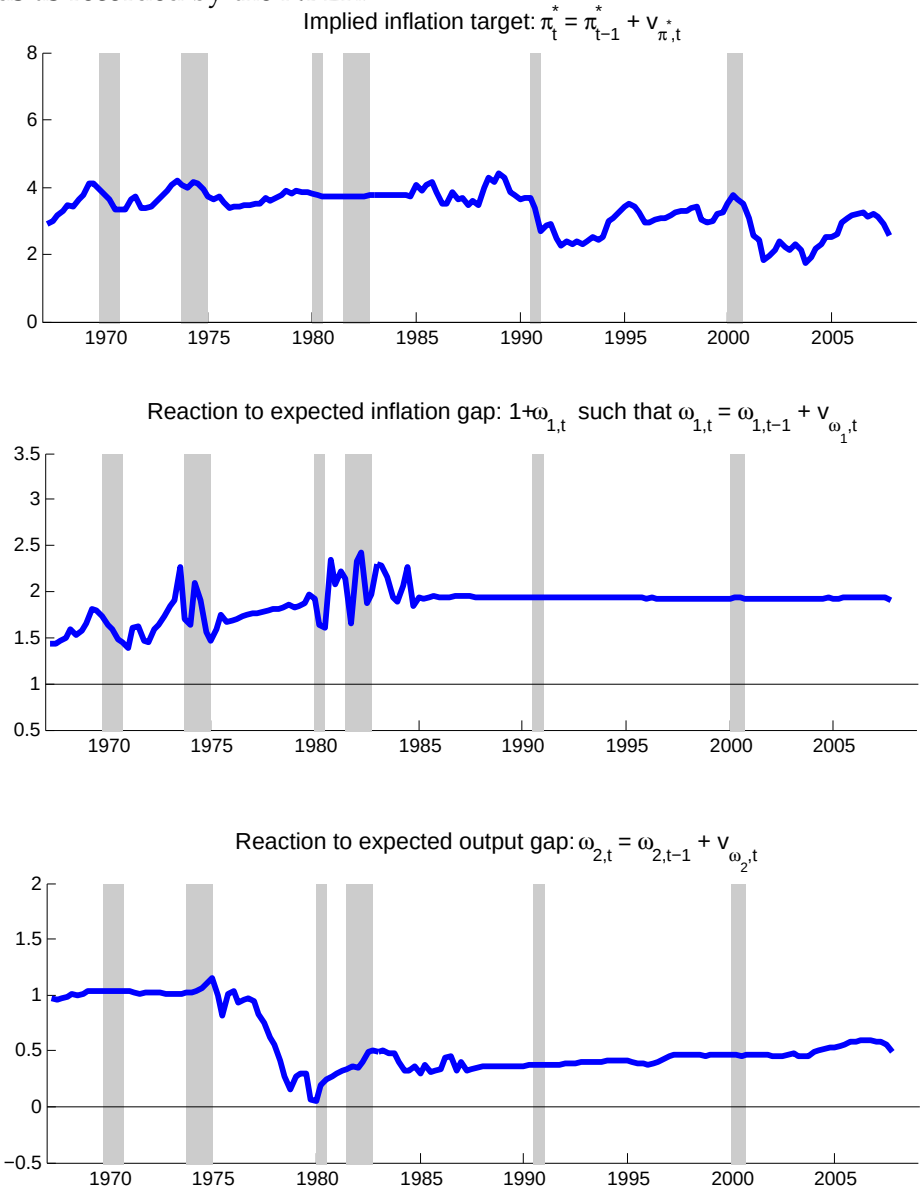


Figure 3.4: Implied time varying structural policy parameters (posterior medians) under identification selection method 2: regime switching between the three best-fit identification strategies under method 1. Structural policy parameters are recovered by feeding the estimated sequences of shocks in Figure 3.3 to the corresponding random walk representation of π_t^*, $\omega_{1,t}$ and $\omega_{2,t}$. Shaded regions denote recession periods as recorded by the NBER.



3.5.5 Robustness Analysis

We now investigate the robustness of our previous results. This exercise is conducted along four dimensions:

• Evaluation of trend inflation estimates.

• Allowing for persistence in the specification of the transitory of component of inflation.

• Changing the break date for second moment parameters.

• Using an alternative measure of real-time output gap.

Specifically, point 1 refers to comparing the measure of trend inflation extracted from M_{R1} (our best-fit model according to method 1) with alternative measures of trend inflation. For point 2, instead of assuming transitory inflation is purely a white noise process, we allow for some persistence in the non-trend component of inflation. In particular we augment the measurement equation for π_t in (3.2.1) by introducing lagged terms of inflation and the policy rate. For point 3, we set the break date for all covariance matrices at 1979Q4 in order to capture the beginning of the Volcker disinflation period. Finally, following Boivin [2006], we re-estimate our models using a measure of real-time expected output gap where the natural rate of unemployment is constructed using an exponential smoothing filter. More precisely, let y_t^* and y_t denote an alternative measure of the natural unemployment rate and the quarterly civilian unemployment rate obtained from the Federal Reserve Bank of St. Louis website, respectively. We then set: $y_t^* = y_{t-1}^* + 0.075(y_t - y_{t-1}^*)$ initialized with $y_0^* = y_1$.

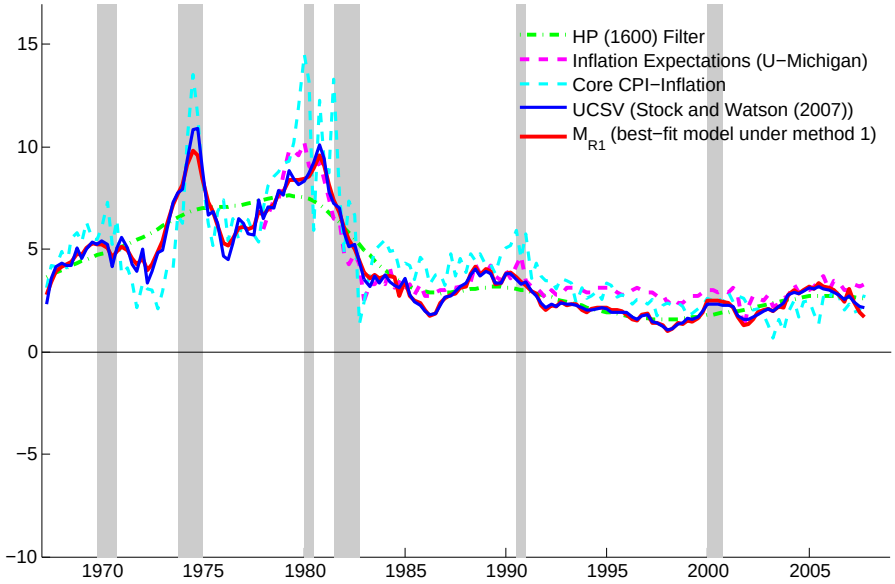
Overall, our results for points 2, 3 and 4 are very similar to what we reported above. Therefore, in what follows we present results for point 1 and report our findings for the other robustness checks in Appendix 3.A.6.

Figure 3.5 illustrates how the measure of trend inflation extracted using our best-fit model under method 1 (M_{R1}) compares with other popular metrics of trend inflation.[38] In particular our measure of trend inflation is closely related to the one obtained using the univariate UC model with stochastic volatility (UCSV) approach

[38]To construct trend inflation, notice that estimation of the system in (3.4.8)-(3.4.9) yields the vector of residuals $\mathbf{e} = \mathbf{Y} - \mathbf{Z}\tilde{\boldsymbol{\alpha}} - \widetilde{\mathbf{Z}}\tilde{\mathbf{v}}$, where $\mathbf{e} = [\mathbf{e}_{\mathfrak{B}}' \ \mathbf{e}_{\mathbf{r}}']'$. Therefore, a vector for trend inflation, $\boldsymbol{\tau}$, can be recovered by collecting the vector of inflation observations, $\boldsymbol{\pi}$, in $\mathbf{Y}$ and setting $\boldsymbol{\tau} = \boldsymbol{\pi} - \mathbf{e}_{\mathfrak{B}}$.

in Stock and Watson [2007].[39] It also correlates well with the survey-based measure of inflation expectations from the University of Michigan.

Figure 3.5: Model-based, non-parametric, survey-based and ad-hoc measures of trend inflation. Shaded regions denote recession periods as recorded by the NBER.



[39] Stock and Watson [2007] construct a measure of trend inflation based on a similar decomposition of inflation as the one presented in (3.2.1). In their paper, however, the volatility of the measurement and state innovations are modeled as state variables following independent random walk processes. To illustrate, given an innovation $e_t \sim \mathcal{N}\left(0,\, \sigma_t^2\right)$, Stock and Watson [2007] define $e_t = \exp\left(\frac{h_t}{2}\right) u_t$ such that u_t follows a standard normal density and $h_t = h_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}\left(0,\, \sigma_\eta^2\right)$. The time series reported in Figure 3.5 corresponding to the UCSV model, is estimated using prior densities calibrated as in Stock and Watson [2007].

3.6 Concluding Remarks

Many approaches for estimation of U.S. trend inflation have recently emerged in the literature. Nonetheless, evidence on what underpins time variation in trend inflation remains scant. In this chapter, we addressed this issue by developing a flexible and yet tightly parameterized econometric framework to empirically assess the role of monetary policy as a driver of trend inflation. In particular, we combined unobserved components modeling of trend inflation with a forward-looking Taylor-type monetary policy rule to construct a bivariate state space model for inflation and the policy rate. As a result, trend inflation and policy rule parameters were treated as latent stochastic factors. This allowed us to statistically evaluate the interdependence between monetary policy and trend inflation in a dynamic setting. Moreover, it allowed us to investigate the relevance of monetary policy as a determinant of trend inflation in comparison to random non-policy factors.

The framework proposed in this chapter is consistent with inflation being affected by both systematic and non-systematic monetary policy. To reconcile longer run policy strategies to its longer run goals, we assumed monetary policy influences on trend inflation are felt through its systematic component. Also, to enable a wider range of identified systematic monetary policy shocks to trend inflation, we postulated several types of parameter instability within a structural representation of the policy rule. Such approach generated particular overidentifying restrictions to our econometric model. Since each overidentifying restriction corresponded to a particular type of systematic monetary policy shock to trend inflation, static and dynamic model selection methods were used to assess the fit of each shock as a candidate driver of trend inflation.

Overall, our results suggest that the increased influence monetary policy exerted over trend inflation during the Great Moderation relative to the Great Inflation reflects a shift in policy emphasis from output gap stabilization towards inflation gap stabilization as well as gradual adjustments in the target inflation rate during both episodes. Nevertheless, while important, monetary policy implications for trend inflation should not be overstated. According to our best-fit models, non-policy shocks accounted for more than 50% of the variation in trend inflation movements during the Great Inflation and Great Moderation.

Finally, in this chapter we did not attempt to uncover the nature of non-policy shocks underpinning trend inflation behaviour. Admittedly, mispecification – inherent to any modeling effort – might have affected some of our estimates, the relative importance of non-policy shocks found here indicates that further work on the identification of such shocks as well as their incorporation into theoretical models of trend

inflation is warranted.

Appendix 3.A Appendix

• *Remark on Notation*:

Throughout this appendix we adopt the following convention to represent entries in a vector: given some vector $\mathbf{x}$, then $\mathbf{x}^{a:b}$ denotes the observations in such vector ranging from entries a to b.

3.A.1 The Gibbs Sampler Used in Section 4 for Method 1

In this appendix we describe the MCMC algorithm used to estimate all models under identification selection Method 1 as described in Section 3.3.1.1. Recall from Section 3.4.2.1 that estimation of such models involve sampling from the following conditional posterior densities:

1. $f(\widetilde{\mathbf{v}}|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\mu}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\widetilde{\mathbf{v}}},\ \overline{\mathbf{D}}_{\widetilde{\mathbf{v}}}\right)$,

2. $f(\boldsymbol{\mu}|\mathbf{Y},\widetilde{\mathbf{v}},\boldsymbol{\theta}_{-\boldsymbol{\mu}}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\boldsymbol{\mu}},\ \overline{\mathbf{D}}_{\boldsymbol{\mu}}\right)$,

3. $f(\Omega_{e,j}|\mathbf{Y},\widetilde{\mathbf{v}},\boldsymbol{\theta}_{-\Omega_{e,j}}) \sim \mathcal{IW}\left(\bar{\nu}_{\Omega_{e,j}}\ \overline{\mathbf{W}}_{\Omega_{e,j}}\right)$ for $j = 1,2$,

4. $f(\sigma^2_{v_{i,j}}|\mathbf{Y},\widetilde{\mathbf{v}},\boldsymbol{\theta}_{-\sigma^2_{v_{i,j}}}) \sim \mathcal{IG}\left(\bar{\nu}_{\sigma^2_{v_{i,j}}}\ \bar{S}_{\sigma^2_{v_{i,j}}}\right)$ for $i = \tau^*, r_0..., r_5$ and $j = 1,2$.

Below we derive expressions for the densities above.

• **Step 1: Sampling $\widetilde{\mathbf{v}}$ (Disturbance Smoothing)**

For convenience the state space representation in (3.4.8) and (3.4.9) is repeated here:

$$\mathbf{Y} = \mathbf{Z}\widetilde{\boldsymbol{\alpha}} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{v}} + \mathbf{e}, \tag{3.A.1}$$

$$\widetilde{\mathbf{v}} \sim \mathcal{N}\left(\mathbf{0},\ \mathbf{L}^{-1}\boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{L}'^{-1}\right), \tag{3.A.2}$$

where $\mathbf{e} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{e}})$ and all matrices and vectors above being defined as in Sections 3.3.1 and 3.4.2. Now let $\mathbf{Y}_{\widetilde{\mathbf{v}}} = \mathbf{Y} - \mathbf{Z}\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}} = \mathbf{L}^{-1}\boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{L}'^{-1}$. Applying Bayes' rule and

standard regression results (see, e.g., Koop et al. (2007)) we have:

$$
\begin{aligned}
f(\widetilde{\mathbf{v}}|\mathbf{Y},\boldsymbol{\theta}) &\propto f(\mathbf{Y}|\widetilde{\mathbf{v}},\boldsymbol{\theta})f(\widetilde{\mathbf{v}}|\boldsymbol{\theta}),\\
f(\widetilde{\mathbf{v}}|\mathbf{Y},\boldsymbol{\theta}) &\propto \exp\left[-\frac{\left(\mathbf{Y}_{\widetilde{\mathbf{v}}}-\widetilde{\mathbf{Z}}\widetilde{\mathbf{v}}\right)'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\left(\mathbf{Y}_{\widetilde{\mathbf{v}}}-\widetilde{\mathbf{Z}}\widetilde{\mathbf{v}}\right)+\widetilde{\mathbf{v}}'\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}\widetilde{\mathbf{v}}}{2}\right],\\
&\propto \exp\left[-\frac{\widetilde{\mathbf{v}}'\left(\widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{Z}}+\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}\right)\widetilde{\mathbf{v}}-2\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{Z}}\widetilde{\mathbf{v}}}{2}\right].
\end{aligned}
$$

The last expression above reveals a Gaussian kernel such that:

$$
\widetilde{\mathbf{v}}|\mathbf{Y},\boldsymbol{\theta}\sim\mathcal{N}\left(\bar{\mathbf{d}}_{\widetilde{\mathbf{v}}},\ \overline{\mathbf{D}}_{\widetilde{\mathbf{v}}}\right),\ \text{where}\ \begin{cases}\bar{\mathbf{d}}_{\widetilde{\mathbf{v}}}=\overline{\mathbf{D}}_{\widetilde{\mathbf{v}}}\widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\mathbf{v}},\\ \overline{\mathbf{D}}_{\widetilde{\mathbf{v}}}=\left(\widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{Z}}+\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}\right)^{-1}.\end{cases} \tag{3.A.3}
$$

Draws from the density above are obtained using the precision sampler in Chan and Jeliazkov [2009]. Also, recall that $\widetilde{\mathbf{v}}=\mathbf{L}^{-1}\mathbf{v}$, therefore, once a draw for $\widetilde{\mathbf{v}}$ is obtained the innovations vector, $\mathbf{v}$, can be easily recovered setting $\mathbf{v}=\mathbf{L}\widetilde{\mathbf{v}}$. We choose to sample $\widetilde{\mathbf{v}}$ to expedite computation as such parameterization allows us to explore the sparse structure of the precision matrix associated with the covariance in (3.A.3), i.e., $\overline{\mathbf{D}}_{\widetilde{\mathbf{v}}}^{-1}=\left(\widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{Z}}+\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}\right)$.[40]

- **Step 2: Sampling** $\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\alpha}\\ \gamma\end{bmatrix}$

First, recall that $\boldsymbol{\alpha}$ is the vector containing initialization conditions and time invariant parameters in the reduced form policy rule, while $\gamma=(\gamma_1,\ \gamma_2)'$ is the vector of loading parameters associated with a monetary policy shock to trend inflation before and after the Great Moderation. In what follows, we show that a simple reparameterization of (3.A.1) allows one to jointly sample $\boldsymbol{\alpha}$ and γ from a Gaussian conditional posterior. For brevity, we use model M_{R1} to illustrate our ideas. Algebraic manipulations for all other models in Table 3.1, however, follow closely the exposition below.

Now, using $\mathbf{Z}$, $\mathbf{R}_\gamma$, $\mathbf{H}_\gamma$ and $\widetilde{\mathbf{v}}$ as defined in Sections 3.3.1 and 3.3.1.1 we can rearrange $\widetilde{\mathbf{Z}}\widetilde{\mathbf{v}}=\mathbf{Z}\mathbf{R}_\gamma\widetilde{\mathbf{v}}$ as:

[40] The interested reader is referred to Chan and Jeliazkov [2009] and Chan [2013] for details.

$$\mathbf{Z}\mathbf{R}_\gamma\widetilde{\mathbf{v}} = \mathbf{Z}\begin{bmatrix} \widetilde{\mathbf{v}}_{\tau^*} + \mathbf{H}_\gamma\widetilde{\mathbf{v}}_{\pi^*} \\ \widetilde{\mathbf{v}}_{\pi^*} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \mathbf{Z}\begin{bmatrix} \widetilde{\mathbf{v}}_{\tau^*} \\ \widetilde{\mathbf{v}}_{\pi^*} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \mathbf{Z}\begin{bmatrix} \mathbf{H}_\gamma\widetilde{\mathbf{v}}_{\pi^*} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$= \mathbf{Z}\begin{bmatrix} \widetilde{\mathbf{v}}_{\tau^*} \\ \widetilde{\mathbf{v}}_{\pi^*} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \mathbf{Z}\begin{bmatrix} \gamma_1\left[\widetilde{\mathbf{v}}'^{1:T_{break}}_{\pi^*} \quad \mathbf{0}'^{T_{break}+1:T}\right]' \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \mathbf{Z}\begin{bmatrix} \gamma_2\left[\mathbf{0}'^{1:T_{break}} \quad \widetilde{\mathbf{v}}'^{T_{break}+1:T}_{\pi^*}\right]' \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$= \begin{bmatrix} \widetilde{\mathbf{v}}_{\tau^*} \\ \widetilde{\mathbf{v}}_{\pi^*} \end{bmatrix} + \begin{bmatrix} \left[\widetilde{\mathbf{v}}'^{1:T_{break}}_{\pi^*} \quad \mathbf{0}'^{T_{break}+1:T}\right]' \\ \mathbf{0} \end{bmatrix}\gamma_1 + \begin{bmatrix} \left[\mathbf{0}'^{1:T_{break}} \quad \widetilde{\mathbf{v}}'^{T_{break}+1:T}_{r_0}\right]' \\ \mathbf{0} \end{bmatrix}\gamma_2,$$

where the first equality uses the structure of $\mathbf{R}_{\mathbf{fl}}\widetilde{\mathbf{v}}$ defined in Section 3.3.1.1 for model M_{R1}. The second and third equalities simply apply the distributive property of matrix multiplication and the fact that $\mathbf{H}_\gamma$ is diagonal.

Next, recall $\widetilde{\boldsymbol{\alpha}} = \mathbf{L}^{-1}\mathbf{F}\boldsymbol{\alpha}$, with $\mathbf{L}$ and $\mathbf{F}$ defined as in Section 3.3.1. Then, letting $\mathbf{Z}\widetilde{\boldsymbol{\alpha}} = \mathbf{Z}\mathbf{L}^{-1}\mathbf{F}\boldsymbol{\alpha} = \mathbf{C}\boldsymbol{\alpha}$, yields an observationally equivalent representation of (3.A.1) as:

$$\mathbf{Y} - \begin{bmatrix} \widetilde{\mathbf{v}}_{\tau^*} \\ \widetilde{\mathbf{v}}_{\pi^*} \end{bmatrix} = \mathbf{C}\boldsymbol{\alpha} + \begin{bmatrix} \left[\widetilde{\mathbf{v}}'^{1:T_{break}}_{\pi^*} \quad \mathbf{0}'^{T_{break}+1:T}\right]' & \left[\mathbf{0}'^{1:T_{break}} \quad \widetilde{\mathbf{v}}'^{T_{break}+1:T}_{\pi^*}\right]' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{\gamma} + \mathbf{e} \iff$$

$$\mathbf{Y}\text{-} = \widetilde{\mathbf{C}}\boldsymbol{\mu} + \mathbf{e}, \tag{3.A.4}$$

where $\mathbf{Y}\text{-} = \mathbf{Y} - \begin{bmatrix} \widetilde{\mathbf{v}}_{\tau^*} \\ \widetilde{\mathbf{v}}_{\pi^*} \end{bmatrix}$ and $\widetilde{\mathbf{C}}$ denotes a matrix concatenation of $\mathbf{C}$ and the matrix preceding $\boldsymbol{\gamma}$, with the column vectors in the latter appended as the last two columns in $\widetilde{\mathbf{C}}$.

Next, remembering we adopt an independent Normal prior for $\boldsymbol{\mu}$, Bayes' rule and

standard regression results (see, e.g.,Koop et al. [2007]) similar to Step 1 yields:

$$\boldsymbol{\mu}|\mathbf{Y},\widetilde{\mathbf{v}},\boldsymbol{\theta}_{-\boldsymbol{\mu}} \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\boldsymbol{\mu}},\ \bar{\mathbf{D}}_{\boldsymbol{\mu}}\right), \text{ where } \begin{cases} \bar{\mathbf{d}}_{\boldsymbol{\mu}} = \bar{\mathbf{D}}_{\boldsymbol{\mu}}\left(\widetilde{\mathbf{C}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}\text{-} + \underline{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}{}^{-1}\underline{\boldsymbol{\mu}}\right), \\ \bar{\mathbf{D}}_{\boldsymbol{\mu}} = \left(\widetilde{\mathbf{C}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{C}} + \underline{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}{}^{-1}\right)^{-1}. \end{cases} \tag{3.A.5}$$

The sampling procedure for $\boldsymbol{\mu}$ is virtually unchanged when allowing for additional regressors such as in the second robustness check discussed in Section 3.5.5. To be precise, let $\mathbf{W}$ denote an exogenous block appropriately adjusted to account for the additional regressors reported in Section 3.5.5 and $\boldsymbol{\delta}$ the vector of time invariant coefficients associated with $\mathbf{W}$. The measurement equation in (3.A.1) can then be expressed as:

$$\mathbf{Y} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{W}\boldsymbol{\delta} + \mathbf{Z}\mathbf{R}_{\boldsymbol{\gamma}}\widetilde{\mathbf{v}} + \mathbf{e}.$$

It is easy to see that the decomposition of $\mathbf{Z}\mathbf{R}_{\boldsymbol{\gamma}}\widetilde{\mathbf{v}}$ shown above is unchanged, leading to a parameterization of the measurement equation as in (3.A.4). The only difference is that the matrix concatenation to form $\widetilde{\mathbf{C}}$ now also requires accounting for $\mathbf{W}$ accordingly and, likewise, setting $\boldsymbol{\mu} = [\boldsymbol{\alpha}\ \boldsymbol{\delta}\ \boldsymbol{\gamma}]'$.

- **Step 3: Sampling $\Omega_{e,1}$ and $\Omega_{e,2}$**

Recall the state space representation in (3.A.1) and (3.A.2) stacks observations as $\mathbf{Y} = \begin{bmatrix} \boldsymbol{\pi} \\ \mathbf{r} \end{bmatrix}$ where $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_T)'$ and $\mathbf{r} = (r_1, \cdots, r_T)'$. As a result, the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{e}}$ is not in the usual block diagonal form which – when combined with a independent inverse wishart prior for $\Omega_{e,1}$ and $\Omega_{e,2}$ – leads to a conditional posterior density which is also a inverse wishart density, hence easy to sample from. Formally we have:

$$\boldsymbol{\Sigma}_{\mathbf{e}} = \mathbb{E}\left[\begin{bmatrix} \mathbf{e}_{\pi} \\ \mathbf{e}_{r} \end{bmatrix}\begin{bmatrix} \mathbf{e}_{\pi}' & \mathbf{e}_{r}' \end{bmatrix}\right] = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}_{\pi}} & \boldsymbol{\Sigma}_{\mathbf{e}_{\pi}\mathbf{e}_{r}} \\ \boldsymbol{\Sigma}_{\mathbf{e}_{\pi}\mathbf{e}_{r}} & \boldsymbol{\Sigma}_{\mathbf{e}_{r}} \end{bmatrix}.$$

where $\boldsymbol{\Sigma}_{\mathbf{e}_{\pi}}$, $\boldsymbol{\Sigma}_{\mathbf{e}_{r}}$ and $\boldsymbol{\Sigma}_{\mathbf{e}_{\pi}\mathbf{e}_{r}}$ are diagonal matrices respectively containing the unconditional variances for $e_{\pi,t}$, $e_{r,t}$ and the covariances between $e_{\pi,t}$ and $e_{r,t}$ before and after the break point, T_{break}. To recast the $\boldsymbol{\Sigma}_{\mathbf{e}}$ in block diagonal form, we define a

permutation matrix, $\mathbf{S}$, with entries given by

$$s_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in \{(1,1),\ (3,2),\ (5,3),\ ,...,(2T-1,T)\}, \\ 1 & \text{if } (i,j) \in \{(2,T+1),\ (4,T+2),\ (6,T+3),\ ,...,(2T,2T)\}, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore:

$$\overbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & & & \vdots \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & & & \\ & & & & & \ddots & & & & & \\ \vdots & & & & & & \ddots & & & & \vdots \\ & & & & & & & & & & \\ & & & & & & & & & & \\ 0 & 0 & \cdots & 1 & & & \cdots & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 & & & \cdots & 0 & \cdots & \cdots & 1 \end{bmatrix}}^{\mathbf{S}} \overbrace{\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \\ \pi_T \\ r_1 \\ r_2 \\ \vdots \\ \\ r_T \end{bmatrix}}^{\mathbf{Y}} = \overbrace{\begin{bmatrix} \pi_1 \\ r_1 \\ \pi_2 \\ r_2 \\ \\ \\ \vdots \\ \\ \pi_T \\ r_T \end{bmatrix}}^{\widetilde{\mathbf{Y}}},$$

and an observationally equivalent representation of the measurement equation in (3.A.1) can, thus, be expressed as follows:

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{A}}_0 \boldsymbol{\alpha} + \widetilde{\mathbf{A}}_1 \widetilde{\mathbf{v}} + \widetilde{\mathbf{e}} \qquad \widetilde{\mathbf{e}} \qquad \sim \mathcal{N}\left(\mathbf{0},\ \widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{e}}}\right), \tag{3.A.6}$$

where:

$$\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{e}}} = \mathbb{E}\left[\begin{bmatrix} \widetilde{\mathbf{e}}_1 \\ \vdots \\ \widetilde{\mathbf{e}}_T \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{e}}_1' & \cdots & \widetilde{\mathbf{e}}_T' \end{bmatrix}\right] = \begin{bmatrix} \mathbf{I}_{T_{break}} \otimes \Omega_{e,1} & \mathbf{0}'_{2(T-T_{break})\times 2T_{break}} \\ \mathbf{0}_{2(T-T_{break})\times 2T_{break}} & \mathbf{I}_{T-T_{break}} \otimes \Omega_{e,2} \end{bmatrix},$$

$\widetilde{\mathbf{Y}} = \mathbf{SY}$, $\widetilde{\mathbf{A}}_0 = \mathbf{SC}$, $\widetilde{\mathbf{A}}_1 = \mathbf{SZR}_\gamma$ and $\widetilde{\mathbf{e}} = \mathbf{Se}$.

Next, by virtue of Bayes' rule combined with standard results in linear regression (see, e.g., Koop et al. [2007]), the conditional posterior for $f(\Omega_{e,j}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\Omega_{e,j}})$ ensues as illustrated below:

- $j = 1$ (sampling pre-break covariance matrix, $\Omega_{e,1}$):

$$
\begin{aligned}
f(\Omega_{e,1}|\widetilde{\mathbf{Y}}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\Omega_{e,1}}) &\propto f(\widetilde{\mathbf{Y}}|\boldsymbol{\theta}, \widetilde{\mathbf{v}}) f(\Omega_{e,1}), \\
&\propto |\Omega_{e,1}^{-1}|^{\frac{T_{break}+\underline{\nu}_{\Omega_{e,1}}-3}{2}} \exp\left[-\frac{\sum_{t=1}^{T_{break}} \widetilde{e}_t' \Omega_{e,1}^{-1} \widetilde{e}_t + \text{trace}\left(\underline{\nu}_{\Omega_{e,1}} \underline{\mathbf{W}}_{\Omega_{e,1}} \Omega_{e,1}^{-1}\right)}{2}\right], \\
&= |\Omega_{e,1}^{-1}|^{\frac{T_{break}+\underline{\nu}_{\Omega_{e,1}}-3}{2}} \exp\left[-\frac{\sum_{t=1}^{T_{break}} \text{trace}\left(\widetilde{e}_t \widetilde{e}_t' \Omega_{e,1}^{-1}\right) + \text{trace}\left(\underline{\nu}_{\Omega_{e,1}} \underline{\mathbf{W}}_{\Omega_{e,1}} \Omega_{e,1}^{-1}\right)}{2}\right], \\
&= |\Omega_{e,1}^{-1}|^{\frac{T_{break}+\underline{\nu}_{\Omega_{e,1}}-3}{2}} \exp\left\{-\frac{\text{trace}\left[\left(\sum_{t=1}^{T_{break}} \widetilde{e}_t \widetilde{e}_t' + \underline{\nu}_{\Omega_{e,1}} \underline{\mathbf{W}}_{\Omega_{e,1}}\right) \Omega_{e,1}^{-1}\right]}{2}\right\}.
\end{aligned}
$$

The last expression above reveals the kernel of a inverse wishart density such that:

$$
\Omega_{e,1}|\widetilde{\mathbf{Y}}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\Omega_{e,1}} \sim \mathcal{IW}\left(\bar{\nu}_{\Omega_{e,1}}\, \overline{\mathbf{W}}_{\Omega_{e,1}}\right), \text{ where } \begin{cases} \bar{\nu}_{\Omega_{e,1}} = T_{break} + \underline{\nu}_{\Omega_{e,1}}, \\ \overline{\mathbf{W}}_{\Omega_{e,1}} = \left(\sum_{t=1}^{T_{break}} \widetilde{e}_t \widetilde{e}_t' + \underline{\nu}_{\Omega_{e,1}} \underline{\mathbf{W}}_{\Omega_{e,1}}\right)^{-1}. \end{cases} \tag{3.A.7}
$$

- $j = 2$ (sampling post-break covariance matrix, $\Omega_{e,2}$):

Applying the same algebraic steps as in the $j = 1$ case yields:

$$
\Omega_{e,2}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\Omega_{e,2}} \sim \mathcal{IW}\left(\bar{\nu}_{\Omega_{e,2}}\, \overline{\mathbf{W}}_{\Omega_{e,2}}\right), \text{ where } \begin{cases} \bar{\nu}_{\Omega_{e,2}} = (T - T_{break}) + \underline{\nu}_{\Omega_{e,2}}, \\ \overline{\mathbf{W}}_{\Omega_{e,2}} = \left(\sum_{t=T_{break}+1}^{T} \widetilde{e}_t \widetilde{e}_t' + \underline{\nu}_{\Omega_{e,2}} \underline{\mathbf{W}}_{\Omega_{e,2}}\right)^{-1}. \end{cases} \tag{3.A.8}
$$

- **Step 4: Sampling $\sigma^2_{\bar{v}_{i,j}}$ for $i = \tau^*, r_0 ..., r_5$ and $j = 1, 2$.**

Lastly, $f(\sigma^2_{\bar{v}_{i,1}}|\mathbf{Y}, \mathbf{v}, \boldsymbol{\theta}_{-\sigma^2_{\bar{v}_{i,1}}})$ and $f(\sigma^2_{\bar{v}_{i,2}}|\mathbf{Y}, \mathbf{v}, \boldsymbol{\theta}_{-\sigma^2_{\bar{v}_{i,2}}})$ are both inverse gamma densities, leading to inverse gamma conditional posteriors after the following algebraic steps:

- $j = 1$ (sampling pre-break variance of state innovations, $\sigma^2_{v_{i,1}}$ for $i = \tau^*, r_0..., r_5$):

$$
\begin{aligned}
f(\sigma^2_{v_{i,1}}|\mathbf{Y}, \mathbf{v}, \boldsymbol{\theta}_{-\sigma^2_{v_{i,1}}}) &\propto f(\mathbf{v}_i|\boldsymbol{\theta}_{-\sigma^2_{v_{i,1}}}) f(\sigma^2_{v_{i,1}}), \\
&\propto {\sigma^2_{v_{i,1}}}^{-\frac{T_{break}}{2}} \exp\left(-\frac{\mathbf{v}_i^{1:T_{break}\prime}\mathbf{v}_i^{1:T_{break}}}{2\sigma^2_{v_{i,1}}}\right) {\sigma^2_{v_{i,1}}}^{-\left(\underline{\nu}_{\sigma^2_{v_{i,1}}}+1\right)} \exp\left(-\frac{\underline{S}_{\sigma^2_{v_{i,1}}}}{\sigma^2_{v_{i,1}}}\right), \\
&\propto {\sigma^2_{v_{i,1}}}^{-\left(\frac{T_{break}}{2}+\underline{\nu}_{\sigma^2_{v_{i,1}}}+1\right)} \exp\left(-\frac{\frac{\sum_{t=1}^{T_{break}} v^2_{i,t}}{2} + \underline{S}_{\sigma^2_{v_{i,1}}}}{\sigma^2_{v_{i,1}}}\right).
\end{aligned}
$$

The expression above reveals the kernel of an inverse gamma density such that:

$$
\sigma^2_{v_{i,1}}|\mathbf{Y}, \mathbf{v}, \boldsymbol{\theta}_{-\sigma^2_{v_{i,1}}} \sim \mathcal{IG}\left(\bar{\nu}_{\sigma^2_{v_{i,1}}}\ \bar{S}_{\sigma^2_{v_{i,1}}}\right), \text{ where } \begin{cases} \bar{\nu}_{\sigma^2_{v_{i,1}}} = \frac{T_{break}}{2} + \underline{\nu}_{\sigma^2_{v_{i,1}}}, \\ \bar{S}_{\sigma^2_{v_{i,1}}} = \frac{\sum_{t=1}^{T_{break}} v^2_{i,t}}{2} + \underline{S}_{\sigma^2_{v_{i,1}}} \end{cases} \quad \text{for } i = \tau^*, r_0..., r_5. \tag{3.A.9}
$$

- $j = 2$ (sampling post-break variance of state innovations, $\sigma^2_{v_{i,2}}$ for $i = \tau^*, r_0..., r_5$):

$$
\sigma^2_{v_{i,2}}|\mathbf{Y}, \mathbf{v}, \boldsymbol{\theta}_{-\sigma^2_{v_{i,2}}} \sim \mathcal{IG}\left(\bar{\nu}_{\sigma^2_{v_{i,2}}}\ \bar{S}_{\sigma^2_{v_{i,j}}}\right), \text{ where } \begin{cases} \bar{\nu}_{\sigma^2_{v_{i,2}}} = \frac{T-T_{break}}{2} + \underline{\nu}_{\sigma^2_{v_{i,2}}}, \\ \bar{S}_{\sigma^2_{v_{i,2}}} = \frac{\sum_{t=T_{break}+1}^{T} v^2_{i,t}}{2} + \underline{S}_{\sigma^2_{v_{i,2}}} \end{cases} \quad \text{for } i = \tau^*, r_0..., r_5. \tag{3.A.10}
$$

A Gibbs sampler form models under method 1 entails sequentially sampling from (3.A.3), (3.A.5), (3.A.7), (3.A.8), (3.A.9) and (3.A.10).

3.A.2 The Augmented Gibbs Sampler to Accommodate Cross-Correlation with Regime-Switching (Method 2)

As discussed in Section 3.4.2.1, when allowing for identification selection with regime switching, the MCMC algorithm discussed above needs to be augmented by the

following two steps:

5. $f(\mathbf{k}|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$,

6. $f(\mathbf{p}_\ell|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\mathbf{p}_\ell}) \sim \mathcal{D}\left(\bar{\alpha}_{\ell,1},\ \bar{\alpha}_{\ell,2},\ \bar{\alpha}_{\ell,3}\right)$ for $\ell = 1, 2, 3$.

• **Step 5: Sampling $\mathbf{k} = (k_1,\ k_2, \cdots,\ k_T)'$**

To sample $\mathbf{k}$, we work with the same parametrization of the measurement equation as in (3.A.6):

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{A}}_0\boldsymbol{\alpha} + \widetilde{\mathbf{A}}_1\widetilde{\mathbf{v}} + \widetilde{\mathbf{e}}.$$

The latter is useful since having the entries in the data vector ordered as $\widetilde{\mathbf{Y}} = (\pi_1,\ r_1, \cdots,\ \pi_T,\ r_T)'$, instead of $\mathbf{Y} = (\pi_1, \cdots, \pi_T,\ r_1, \cdots, r_T)'$, as in equation (3.A.1), ensures that the filtering step in the recursive process described below is correctly conducted. To be precise, sampling $\mathbf{k}$ requires the evaluation of the filtering distribution $f(k_t|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ for $t = 1, ..., T$, hence conditional on information for both inflation and the federal funds rate available up to time t, i.e., $\widetilde{\mathbf{Y}}^{1:t} = (\widetilde{\mathbf{y}}_1,\ \cdots,\ \widetilde{\mathbf{y}}_t)' = (\overbrace{\pi_1,\ r_1}^{\widetilde{\mathbf{y}}_1}, \cdots, \overbrace{\pi_t,\ r_t}^{\widetilde{\mathbf{y}}_t})'$. In contrast, if one works with the parametrization in (3.A.1), the ordering of entries in $\mathbf{Y}$ would lead to an incorrect sampling posterior probability for $\mathbf{k}$, since the filtering distribution being recursively evaluated for $t = 1, ..., T$, would be $f(k_t|\mathbf{Y}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ instead of $f(k_t|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$. To be clear, notice that when $t = 1$ the conditional factor in $f(k_t|\mathbf{Y}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ accounts for the first two observations of inflation, i.e, $(\pi_1,\ \pi_2)'$ instead of $(\pi_1\ r_1)'$, since $\mathbf{Y}^{1:T} = (\mathbf{y}_1, \cdots, \mathbf{y}_T)' = (\overbrace{\pi_1,\ \pi_2}^{\mathbf{y}_1}, \cdots, \overbrace{r_{T-1},\ r_T}^{\mathbf{y}_T})'$.

In what follows we provide a condensed discussion of the two-sided filter proposed by Chib [1996], which we adopt here to sample $\mathbf{k} = (k_1,\ k_2, \cdots,\ k_T)'$. The reader is referred to such paper for a comprehensive description of the algorithm. Specifically, below we present the distributions one needs to construct to conduct the predicting and updating steps of the filter. In particular, Chib [1996] suggests a multi-move sampling scheme for $\mathbf{k}$, which entails sequentially sampling each discrete state, $k_t \in \{1,\ ,2,\ 3\}$, from the following product of multinomial distributions:

$$f(\mathbf{k}|\widetilde{\mathbf{Y}}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) = f(k_T|\widetilde{\mathbf{Y}}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) \times \cdots \times f(k_t|\widetilde{\mathbf{Y}}, \mathbf{k}^{t+1:T}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) \times \cdots \times f(k_1|\widetilde{\mathbf{Y}}, \mathbf{k}^{2:T}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}). \tag{3.A.11}$$

Notably, all probability distributions above, with the exception of the first one, take the form $f(k_t|\mathbf{Y}, \mathbf{k}^{t+1:T}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ for $t = 1, ..., T-1$. Chib [1996] then shows that by

applying Bayes' rule plus some algebraic manipulations yield:

$$f(k_t|\widetilde{\mathbf{Y}}, \mathbf{k}^{t+1:T}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) = \frac{f(k_t|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) f(k_{t+1}|k_t, \boldsymbol{\theta})}{\sum_{\ell=1}^{3} f(k_t = \ell|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) f(k_{t+1}|k_t = \ell, \boldsymbol{\theta})}. \tag{3.A.12}$$

Now, note that sampling k_t is simply a matter of sampling from a multinomial distribution with success probability given by the expression above. Evaluating the multinomial predictive prior, $f(k_{t+1}|k_t, \boldsymbol{\theta})$, is straightforward. However, one still needs to evaluate the conditional posterior, $f(k_t|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$, which marginalizes over past and future discrete states.

Again, applying Bayes' rule and rearranging terms, yield:

$$f(k_t|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) = \frac{f(k_t|\widetilde{\mathbf{Y}}^{1:t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) f(\widetilde{\mathbf{y}}_t|\widetilde{\mathbf{Y}}^{1:t-1}, k_t, \widetilde{\mathbf{v}}, \boldsymbol{\theta})}{\sum_{\ell=1}^{3} f(k_t = \ell|\widetilde{\mathbf{Y}}^{1:t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) f(\widetilde{\mathbf{y}}_t|\widetilde{\mathbf{Y}}^{1:t-1}, k_t = \ell, \widetilde{\mathbf{v}}, \boldsymbol{\theta})}. \tag{3.A.13}$$

The second density in the expression above is simply a Gaussian component of the joint likelihood $f(\widetilde{\mathbf{Y}}|\boldsymbol{\theta}, \mathbf{k}, \widetilde{\mathbf{v}})$ and therefore can be readily evaluated. To evaluate $f(k_t|\widetilde{\mathbf{Y}}^{1:t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$, note that by the law of total probability and using the property of first-order Markovian processes which ensures $f(k_t|\widetilde{\mathbf{Y}}^{1:t-1}, k_{t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) = f(k_t|k_{t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ we have:

$$f(k_t|\widetilde{\mathbf{Y}}^{1:t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) = \sum_{\ell=1}^{3} f(k_t|k_{t-1} = \ell, \widetilde{\mathbf{v}}, \boldsymbol{\theta}) f(k_{t-1} = \ell|\widetilde{\mathbf{Y}}^{1:t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}). \tag{3.A.14}$$

The mechanics of the algorithm for sampling $\mathbf{k}$ can, thus, be summarized as follows:

At each iteration of the MCMC sampler and beginning at $t = 1$ for $t = 1, ..., T$:

Step 5.1: evaluate predictive distribution in (3.A.14) to obtain a 1×3 vector containing the values of $f(k_t = \ell|\widetilde{\mathbf{Y}}^{1:t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ for $\ell = 1, 2, 3$;[41]

Step 5.2: use the values obtained from the previous step to compute the updating distribution in (3.A.13). In other words, evaluate (3.A.13) for $\ell = 1, 2, 3$ to construct a 1×3 vector containing the values of $f(k_t = \ell|\widetilde{\mathbf{Y}}^{1:t}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$;

Step 5.3: when $t = T$ the previous step gives the first distribution in the right-hand side of (3.A.11). Use (3.A.13) to simulate k_T. Such draw initiates the iterative sampling procedure described in (3.A.11), which in turn yields draws for the remain-

[41] At $t = 1$ we initialize $f(k_t|k_{t-1}, \widetilde{\mathbf{v}}, \boldsymbol{\theta})$ by setting it to be the stationary distribution of the chain. That is, a transition matrix $P(k_0)$ such that $\boldsymbol{\omega}_0 P(k_0) = \lambda \boldsymbol{\omega}_0$, where $\boldsymbol{\omega}_0$ is the left eigenvector corresponding to the eigenvalue $\lambda = 1$.

ing states in $\mathbf{k}^{1:T-1}$. In particular, each element in $\mathbf{k}^{1:T-1}$ is obtained by drawing from a multinomial distribution with success probability (i.e. the probability associated with a particular regime, $\ell = 1, 2, 3$) defined in (3.A.12).

• **Step 6: Sampling the transition probability P**

Since the Dirichlet prior adopted for the rows in the transition matrix, $\mathbf{P}$, in 3.4.1 is a conjugate one, then the conditional posterior, $f(\mathbf{p}_\ell|\mathbf{Y}, \widetilde{\mathbf{v}}, \boldsymbol{\theta}_{-\mathbf{p}_\ell})$ for each row vector $\mathbf{p}_i$ for $i = 1, 2, 3$ is also a Dirichlet distribution. As a result, Dirichlet draws for each transition probability from state ℓ_{old} to state ℓ for $\ell_{old}, \ell = 1, 2, 3$, i.e., $p_{\ell_{old},\ell}$ can be obtained as follows:

For any row vector, say, $\mathbf{p_1}$, each entry can be simulated by letting

$$p_{\ell_{old},1} = \frac{x_1}{\sum_{\ell=1}^{3} x_\ell}, \qquad x_\ell \sim \mathcal{G}\left(\underline{\varsigma}_{\ell_{old},1} + n_{\ell_{old},1},\ 1\right),$$

where $\mathcal{G}$ is a Gamma density, $\underline{\varsigma}_{\ell_{old},1}$ is the prior parameter given in Section 3.4.1 for the Dirichlet distribution corresponding to entry $\ell_{old}, 1$ in the row vector $\mathbf{p}_1$ in (3.4.1) and $n_{\ell_{old},1}$ is the total number of one-step transitions sampled from state ℓ_{old} to state 1.

3.A.3 Exclusion Restriction Matrix ($\mathbf{R}_\gamma$) for Each Identification Strategy of a Policy Shock to Trend Inflation

Below we present the structure of the exclusion restrictions matrix, $\mathbf{R}_\gamma$, for all models in Table 3.1:

• M_{R1} :

$$\mathbf{R}_\gamma = \begin{bmatrix} \mathbf{I}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & -\mathbf{I}_T & \mathbf{0}_T & \cdots & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & \ddots & & \mathbf{0}_T & & \vdots \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}$$

- M_{R2} :

$$\mathbf{R}_\gamma = \begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & -\mathbf{I}_T & \cdots & & \mathbf{0}_T \\ \vdots & \ddots & \mathbf{I}_T & & & \vdots \\ & & & \mathbf{0}_T & & \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}$$

- M_{R3} :

$$\mathbf{R}_\gamma = \begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & \cdots & & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & & \ddots & \mathbf{I}_T & & \vdots \\ & & & & \mathbf{0}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}$$

- M_{R4} :

$$\mathbf{R}_\gamma = \begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T & \mathbf{H}_\gamma & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{0}_T & \cdots & & & \mathbf{0}_T \\ & & \mathbf{0}_T & & & \\ \vdots & & \ddots & \mathbf{0}_T & & \vdots \\ & & & & \mathbf{I}_T & \\ \mathbf{0}_T & \cdots & & & & \mathbf{0}_T \end{bmatrix}$$

- M_{R5} :

$$\mathbf{R}_\gamma = \begin{bmatrix} \mathbf{I}_T & \mathbf{H}_\gamma & \mathbf{0}_T & \cdots & & \mathbf{0}_T \\ \mathbf{0}_T & -\mathbf{I}_T & \mathbf{0}_T & \cdots & & \mathbf{0}_T \\ \vdots & \mathbf{I}_T & \mathbf{0}_T & & & \vdots \\ & \mathbf{I}_T & \ddots & \ddots & & \\ & \mathbf{I}_T & & & & \\ \mathbf{0}_T & \mathbf{0}_T & \cdots & & \mathbf{0}_T & \mathbf{0}_T \end{bmatrix}$$

• M_{R6} :

$$\mathbf{R}_\gamma = \begin{bmatrix} \mathbf{I}_T & \mathbf{0}_T & \mathbf{0}_T & \cdots & \mathbf{H}_\gamma \\ \mathbf{0}_T & \vdots & \mathbf{0}_T & \cdots & -\mathbf{I}_T \\ \vdots & & \mathbf{0}_T & & -\mathbf{I}_T \\ & & \ddots & \ddots & -\mathbf{I}_T \\ & & & & -\mathbf{I}_T \\ \mathbf{0}_T & \mathbf{0}_T & \cdots & & \mathbf{I}_T \end{bmatrix}$$

3.A.4 Derivation of the Integrated Likelihood Used for Model Comparison Under Static Identification Selection

In this appendix we provide details on the computation of the metrics used for model comparison reported in Table 1: DIC, $\mathbb{E}$ (log likelihood), BIC and log marginal likelihood. In particular, all such metrics are based on a variant of the likelihood function where the state innovation vector, $\widetilde{\mathbf{v}}$, is integrated out, hence making the parameters in $\boldsymbol{\theta}$ the only conditional factor in the likelihood expression. As pointed out in Chan and Grant [2014], such strategy allows one to simulate the likelihood function with smaller numerical standard errors and consequently delivering more robust model comparison metrics.

Now, recall that when discussing the MCMC sampler in Appendix 3.A.1 we worked with the following state space representation of the model:

$$\begin{aligned} \mathbf{Y} &= \mathbf{Z}\widetilde{\boldsymbol{\alpha}} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{v}} + \mathbf{e}, \\ \widetilde{\mathbf{v}} &\sim \mathcal{N}\left(\mathbf{0},\ \mathbf{L}^{-1}\boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{L}'^{-1}\right). \end{aligned}$$

The expression of the measurement equation above indicates that to integrate $\widetilde{\mathbf{v}}$ from the likelihood function, $f(\mathbf{Y}|\widetilde{\mathbf{v}}, \boldsymbol{\theta})$, one needs to derive an expression for the integrated likelihood function, $f(\mathbf{Y}|\boldsymbol{\theta})$. Below, we describe a strategy to do that which is an alternative to the usual Kalman filter approach and allows for fast computation of $f(\mathbf{Y}|\boldsymbol{\theta})$ by exploring the sparse structure of the matrices in our models. First, recall $\mathbf{Y}$ and $\widetilde{\mathbf{v}}$ are $2T \times 1$ and $6T \times 1$ vectors respectively and let $c_1 = \frac{1}{(2\pi)^{\frac{2T}{2}}}|\boldsymbol{\Sigma}_{\mathbf{e}}|^{-\frac{1}{2}}$, $c_2 =$

$\frac{1}{(2\pi)^{\frac{6T}{2}}}|\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}|^{-\frac{1}{2}}$, $\mathbf{Y}_{\widetilde{\mathbf{v}}} = \mathbf{Y} - \mathbf{Z}\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}} = \mathbf{L}^{-1}\boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{L}'^{-1}$. Then, $f(\mathbf{Y}|\boldsymbol{\theta})$ is given by:

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}) &= \int f(\mathbf{Y}|\widetilde{\mathbf{v}}, \boldsymbol{\theta}) f(\widetilde{\mathbf{v}}|\boldsymbol{\theta}) d\widetilde{\mathbf{v}}, \\ &= c_1 c_2 \int \exp\left[-\frac{\left(\mathbf{Y}_{\widetilde{\mathbf{v}}} - \widetilde{\mathbf{Z}}\widetilde{\mathbf{v}}\right)' \boldsymbol{\Sigma}_{\mathbf{e}}^{-1} \left(\mathbf{Y}_{\widetilde{\mathbf{v}}} - \widetilde{\mathbf{Z}}\widetilde{\mathbf{v}}\right)}{2}\right] \exp\left(\frac{\widetilde{\mathbf{v}}'\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}\widetilde{\mathbf{v}}}{2}\right) d\widetilde{\mathbf{v}}, \\ &= c_1 c_2 \exp\left(-\frac{\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}}}{2}\right) \int \exp\left[-\frac{\widetilde{\mathbf{v}}'\left(\widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{Z}} + \widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}\right)\widetilde{\mathbf{v}} - 2\widetilde{\mathbf{v}}'\widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}}}{2}\right] d\widetilde{\mathbf{v}}. \end{aligned} \tag{3.A.15}$$

Now, let $\mathbf{F} = \widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\widetilde{\mathbf{Z}} + \widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}^{-1}$ and $\mathbf{f} = \widetilde{\mathbf{Z}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}}$. Therefore, by completing the square we have:

$$\begin{aligned} \widetilde{\mathbf{v}}'\mathbf{F}\widetilde{\mathbf{v}} - 2\widetilde{\mathbf{v}}'\mathbf{f} &= \widetilde{\mathbf{v}}'\mathbf{F}\widetilde{\mathbf{v}} - 2\widetilde{\mathbf{v}}'\mathbf{f} + \mathbf{f}'\mathbf{F}^{-1}\mathbf{f} - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}, \\ &= \left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)' \mathbf{F} \left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right) - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}. \end{aligned}$$

Plugging the result above into (3.A.15) then yields:

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}) &= c_1 c_2 \exp\left(-\frac{\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}}}{2}\right) \int \exp\left[-\frac{\left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)' \mathbf{F} \left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right) - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}}{2}\right] d\widetilde{\mathbf{v}}, \\ &= c_1 c_2 \exp\left(-\frac{\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}} - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}}{2}\right) \int \exp\left[-\frac{\left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)' \mathbf{F} \left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)}{2}\right] d\widetilde{\mathbf{v}}. \end{aligned} \tag{3.A.16}$$

Next, notice that the integrand in the expression above reveals a Gaussian kernel for $\widetilde{\mathbf{v}}$. Hence, for some normalizing constant, c_3, it follows that:

$$\begin{aligned} &\int c_3 \exp\left[-\frac{\left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)' \mathbf{F} \left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)}{2}\right] d\widetilde{\mathbf{v}} = 1, \\ &\implies \int \exp\left[-\frac{\left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)' \mathbf{F} \left(\widetilde{\mathbf{v}} - \mathbf{F}^{-1}\mathbf{f}\right)}{2}\right] d\widetilde{\mathbf{v}} = \frac{1}{c_3}, \end{aligned} \tag{3.A.17}$$

where $c_3 = \frac{1}{(2\pi)^{\frac{6T}{2}}}|\mathbf{F}|^{\frac{1}{2}}$. Finally, using the result from (3.A.17) in (3.A.16) gives:

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}) &= \frac{c_1 c_2}{c_3}\exp\left(-\frac{\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}} - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}}{2}\right), \\ &= (2\pi)^{-T}|\mathbf{F}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_{\mathbf{e}}|^{-\frac{1}{2}}|\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}|^{-\frac{1}{2}}\exp\left(-\frac{\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}} - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}}{2}\right), \end{aligned}$$

which implies that the integrated log-likelihood used in all metrics of model comparison in Table 1 is computed as follows:

$$\log f(\mathbf{Y}|\boldsymbol{\theta}) = -T\log(2\pi) - \frac{1}{2}\left(\log|\mathbf{F}| + \log|\boldsymbol{\Sigma}_{\mathbf{e}}| + \log|\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{v}}}|\right) - \left(\frac{\mathbf{Y}_{\widetilde{\mathbf{v}}}'\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}\mathbf{Y}_{\widetilde{\mathbf{v}}} - \mathbf{f}'\mathbf{F}^{-1}\mathbf{f}}{2}\right).$$

Now let $size(\boldsymbol{\theta})$ denote the number of parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{\text{mode}}$ the maximum a posteriori estimate which maximizes the integrated likelihood, $f(\mathbf{Y}|\boldsymbol{\theta}_{\text{mode}})$. The expression above is then used to construct the DIC, $\mathbb{E}(\log\text{ likelihood})$, BIC and log marginal likelihood as follows:

$$\begin{aligned} \text{DIC} &= -4\sum_{l=1}^{L}\log f(\mathbf{Y}|\boldsymbol{\theta}^l) + 2\log f(\mathbf{Y}|\boldsymbol{\theta}_{\text{mode}}), \\ \mathbb{E}(\log\text{ likelihood}) &= \sum_{l=1}^{L}\log f(\mathbf{Y}|\boldsymbol{\theta}^l), \\ \text{BIC} &= \log f(\mathbf{Y}|\boldsymbol{\theta}_{\text{mode}}) - \frac{size(\boldsymbol{\theta})}{2}\log(2T), \\ \log(\text{marginal likelihood}) &= \log\left(\frac{1}{L}\sum_{l=1}^{L}\frac{g(\boldsymbol{\theta}^l)}{f(\mathbf{Y}|\boldsymbol{\theta}^l)f(\boldsymbol{\theta}^l)}\right)^{-1}. \end{aligned}$$

To compute the log marginal likelihood using the Gelfand and Dey [1994] approach illustrated above we follow Geweke [1999] and define for some $p \in (0,\ 1)$:

$$g(\boldsymbol{\theta}^l) = p^{-1}(2\pi)^{-\frac{size(\boldsymbol{\theta})}{2}}\left|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^l\right|^{-1}\exp\left[-\frac{\left(\boldsymbol{\theta}^l - \widehat{\boldsymbol{\theta}}\right)'\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^l\right)^{-1}\left(\boldsymbol{\theta}^l - \widehat{\boldsymbol{\theta}}\right)}{2}\right],$$

where $\widehat{\boldsymbol{\theta}}$ is constructed using the average of draws for $\boldsymbol{\theta}$ obtained from the Gibbs sampler and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^l$ denotes approximation of the posterior covariance matrix given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^l = \frac{1}{L}\sum_{l=1}^{L}\left(\boldsymbol{\theta}^l - \widehat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}^l - \widehat{\boldsymbol{\theta}}\right)'.$$

Also, the support of $g(\boldsymbol{\theta}^l)$ is truncated such that Gibbs draws for $\boldsymbol{\theta}^l$ is accepted only if:

$$\left(\boldsymbol{\theta}^l - \widehat{\boldsymbol{\theta}}\right)' \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^l\right)^{-1} \left(\boldsymbol{\theta}^l - \widehat{\boldsymbol{\theta}}\right) \leq \chi_p^2(size(\boldsymbol{\theta})).$$

In other words, to construct the log marginal likelihood we only use posterior draws of $\boldsymbol{\theta}^l$ which ensure that the quadratic form above yields values which fall within the p^{th} percentile of a chi-square distribution with $size(\boldsymbol{\theta})$ degrees of freedom. In this chapter we set $p = 99$th percentile.

3.A.5 Posterior Estimates

Table 3.5: Prior and posterior statistics for the three best-fit models under method 1 (time invariant cross-correlation)

M_{R1} - Inflation target adjustment shock model				
			Prior	Posterior
Parameter	Description	Density	Mean (sd)	Mean (sd)
τ_0	Initial condition for trend inflation	Normal(2.02, 1)	2.02 (1.00)	2.38 (0.71)
$\beta_{0,0}$	Initial condition for drifting intercept in the policy rule	Normal(0.25, 1)	0.25 (1.00)	0.90 (0.65)
β_1	Inflation expectation coefficient	Normal(0.75, 1)	0.75 (1.00)	0.29 (0.09)
β_2	Output gap expectation coefficient	Normal(0.25, 1)	0.25 (1.00)	0.19 (0.10)
β_3	Money growth expectation coefficient	Normal(0.50, 1)	0.50 (1.00)	-0.04 (0.02)
β_4	Policy smoothing coefficient	Normal(0.50, 1)	0.50 (1.00)	0.69 (0.05)
γ_1	Transmission of a systematic policy shock to τ_t during the Great Inflation	Normal(0.00, 1)	0.00 (1.00)	0.93 (0.69)
γ_2	Transmission of a systematic policy shock to τ_t during the Great Moderation	Normal(0.00, 1)	0.00 (1.00)	0.63 (0.30)
$\sigma^2_{\tau,1}$			10.0 (3.16)	0.74 (0.22)
$\sigma^2_{r,1}$	Second moment parameters in $\Omega_{e,1}$ during the Great Inflation	Inverse Wishart(I_2, 10)	10.0 (3.16)	1.26 (0.24)
$\sigma_{\pi r,1}$			0.00 (3.16)	-0.15 (0.16)
$\sigma^2_{\tau,2}$			3.20 (1.01)	0.23 (0.05)
$\sigma^2_{r,2}$	Second moment parameters in $\Omega_{e,2}$ during the Great Moderation	Inverse Wishart($0.32I_2$, 10)	3.20 (1.01)	0.06 (0.01)
$\sigma_{\pi r,2}$			0.00 (1.01)	-0.02 (0.02)
$\sigma^2_{v_{\tau^*},1}$	Variance of non-policy shock to trend inflation during the Great Inflation	Inverse Gamma(10, 2.25)	0.25 (0.09)	0.46 (0.18)
$\sigma^2_{v_{\pi^*},1}$	Variance of inflation target adjustment shock during the Great Inflation	Inverse Gamma(10, 2.25)	0.25 (0.09)	0.16 (0.03)
$\sigma^2_{v_{\tau^*},2}$	Variance of non-policy shock to trend inflation during the Great Moderation	Inverse Gamma(10, 0.09)	0.03 (0.004)	0.11 (0.03)
$\sigma^2_{v_{\pi^*},2}$	Variance of inflation target adjustment shock during the Great Moderation	Inverse Gamma(10, 0.09)	0.03 (0.004)	0.09 (0.02)
M_{R2} - Inflation gap stabilization shock model				
			Prior	Posterior
Parameter	Description	Density	Mean (sd)	Mean (sd)
τ_0	Initial condition for trend inflation	Normal(2.02, 1)	2.02 (1.00)	2.45 (0.70)
$\beta_{0,0}$	Initial condition for drifting intercept in the policy rule	Normal(0.25, 1)	0.25 (1.00)	0.40 (0.41)
$\beta_{1,0}$	Initial condition for drifting coefficient for inflation expectation	Normal(0.75, 1)	0.75 (1.00)	0.01 (0.38)
β_2	Output gap expectation coefficient	Normal(0.25, 1)	0.25 (1.00)	0.26 (0.10)
β_3	Money growth expectation coefficient	Normal(0.50, 1)	0.50 (1.00)	-0.03 (0.03)
β_4	Policy smoothing coefficient in the policy rule	Normal(0.50, 1)	0.50 (1.00)	0.85 (0.04)
γ_1	Transmission of a systematic policy shock to τ_t during the Great Inflation	Normal(0.00, 1)	0.00 (1.00)	0.07 (0.64)
γ_2	Transmission of a systematic policy shock to τ_t during the Great Moderation	Normal(0.00, 1)	0.00 (1.00)	0.49 (0.36)
$\sigma^2_{\tau,1}$			10.0 (3.16)	0.76 (0.23)
$\sigma^2_{r,1}$	Second moment parameters in during the Great Inflation	Inverse Wishart(I_2, 10)	10.0 (3.16)	0.45 (0.14)
$\sigma_{\pi r,1}$			0.00 (3.16)	-0.17 (0.16)
$\sigma^2_{\tau,2}$			3.20 (1.01)	0.25 (0.06)
$\sigma^2_{r,2}$	Second moment parameters in during the Great Moderation	Inverse Wishart($0.32I_2$, 10)	3.20 (1.01)	0.05 (0.01)
$\sigma_{\pi r,2}$			0.00 (1.01)	-0.02 (0.02)
$\sigma^2_{v_{\tau^*},1}$	Variance of non-policy shock to trend inflation during the Great Inflation	Inverse Gamma(10, 2.25)	0.25 (0.09)	0.58 (0.18)
$\sigma^2_{v_{\omega_1},1}$	Variance of inflation gap stabilization shock during the Great Inflation	Inverse Gamma(10, 2.25)	0.25 (0.09)	0.12 (0.02)
$\sigma^2_{v_{\tau^*},2}$	Variance of non-policy shock to trend inflation during the Great Moderation	Inverse Gamma(10, 0.09)	0.03 (0.004)	0.10 (0.02)
$\sigma^2_{v_{\omega_1},2}$	Variance of inflation gap stabilization shock during the Great Moderation	Inverse Gamma(10, 0.09)	0.03 (0.004)	0.06 (0.01)
M_{R3} - Output gap stabilization shock model				
			Prior	Posterior
Parameter	Description	Density	Mean (sd)	Mean (sd)
τ_0	Initialization of trend inflation	Normal(2.02, 1)	2.02 (1.00)	2.50 (0.71)
β_0	Intercept in the policy rule	Normal(0.25, 1)	0.25 (1.00)	0.21 (0.15)
β_1	Inflation expectation coefficient	Normal(0.75, 1)	0.75 (1.00)	0.43 (0.07)
$\beta_{2,0}$	Initialization of drifting coefficient for the output gap expectation	Normal(0.25, 1)	0.25 (1.00)	0.01 (0.51)
β_3	Money growth expectation coefficient	Normal(0.50, 1)	0.50 (1.00)	-0.08 (0.01)
β_4	Policy smoothing coefficient in the policy rule	Normal(0.50, 1)	0.50 (1.00)	0.83 (0.03)
γ_1	Transmission of a systematic policy shock to τ_t during the Great Inflation	Normal(0.00, 1)	0.00 (1.00)	0.76 (0.83)
γ_2	Transmission of a systematic policy shock to τ_t during the Great Moderation	Normal(0.00, 1)	0.00 (1.00)	0.12 (0.38)
$\sigma^2_{\tau,1}$			10.0 (3.16)	0.70 (0.22)
$\sigma^2_{r,1}$	Second moment parameters in $\Omega_{e,1}$ during the Great Inflation	Inverse Wishart(I_2, 10)	10.0 (3.16)	0.81 (0.20)
$\sigma_{\pi r,1}$			0.00 (3.16)	-0.01 (0.19)
$\sigma^2_{\tau,2}$			3.20 (1.01)	0.24 (0.06)
$\sigma^2_{r,2}$	Second moment parameters in $\Omega_{e,2}$ during the Great Moderation	Inverse Wishart($0.32I_2$, 10)	3.20 (1.01)	0.09 (0.01)
$\sigma_{\pi r,2}$			0.00 (1.01)	-0.001 (0.02)
$\sigma^2_{v_{\tau^*},1}$	Variance of non-policy shock to trend inflation during the Great Inflation	Inverse Gamma(10, 2.25)	0.25 (0.09)	0.56 (0.18)
$\sigma^2_{v_{\omega_2},1}$	Variance of output gap stabilization shock during the Great Inflation	Inverse Gamma(10, 2.25)	0.25 (0.09)	0.16 (0.03)
$\sigma^2_{v_{\tau^*},2}$	Variance of non-policy shock to trend inflation during the Great Moderation	Inverse Gamma(10, 0.09)	0.03 (0.004)	0.12 (0.03)
$\sigma^2_{v_{\omega_2},2}$	Variance of output gap stabilization shock during the Great Moderation	Inverse Gamma(10, 0.09)	0.03 (0.004)	0.09 (0.02)

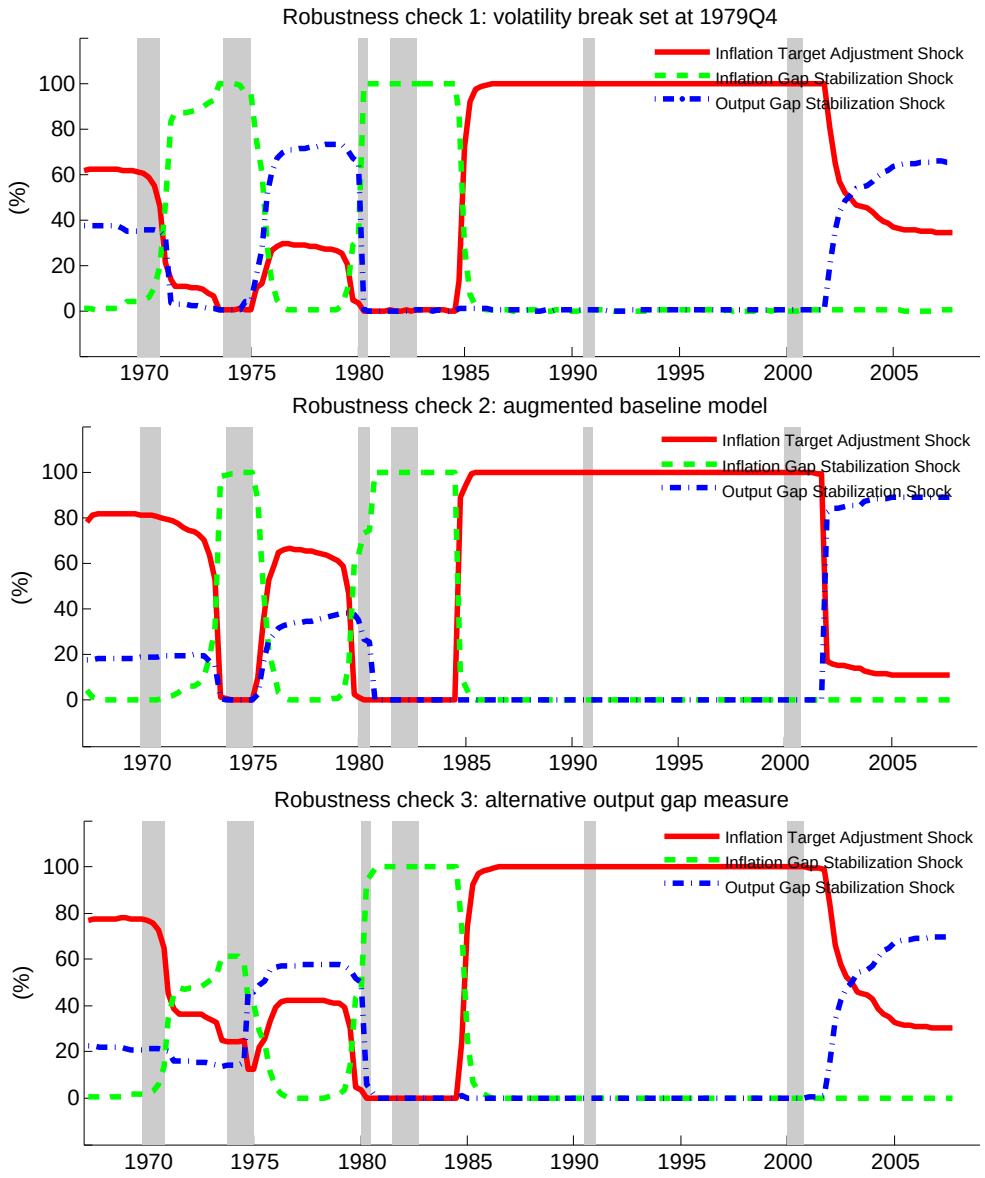
3.A.6 Robustness Checks: Results

In this appendix we report robustness results for points 2, 3 and 4 discussed in Section 3.5.5.

Table 3.6: Variance decomposition results for policy shocks to trend inflation under robustness checks 2, 3 and 4 in Section 3.5.5. Results reported below are for the three best-fit models under identification selection method 1. Baseline specification refers to the model as presented in Section 3.2.

Robustness check 1: volatility break set at 1979Q4		
Model	Great Inflation	Great Moderation
M_{R1}	31.27%	48.12%
M_{R2}	9.55%	19.71%
M_{R3}	13.44%	49.12%
Robustness check 2: augmented baseline specification		
Model	Great Inflation	Great Moderation
M_{R1}	32.11%	44.31%
M_{R2}	9.11%	17.90%
M_{R3}	9.10%	48.12%
Robustness check 3: alternative output gap measure		
Model	Great Inflation	Great Moderation
M_{R1}	30.47%	47.25%
M_{R2}	8.83%	18.25%
M_{R3}	11.25%	47.39%
Baseline specification		
Model	Great Inflation	Great Moderation
M_{R1}	31.74%	48.87%
M_{R2}	8.33%	18.50%
M_{R3}	28.31%	49.81%

Figure 3.6: Posterior probabilities (posterior medians) for each type of systematic monetary policy shock driving trend inflation under method 2: regime switching between the three best-fit identification strategies under method 1. Shaded regions denote recession periods as recorded by the NBER.



3.A.7 MCMC Diagnostics

In this appendix we report inefficiency factors of the posterior draws for parameters and innovations sample using methods 1 and 2. For the sake of brevity, for method 1 we only report results for the best-fit model, M_{R1}. Inefficiency factors are calculated using a common metric (see, e.g. Chib [2001]) given by:

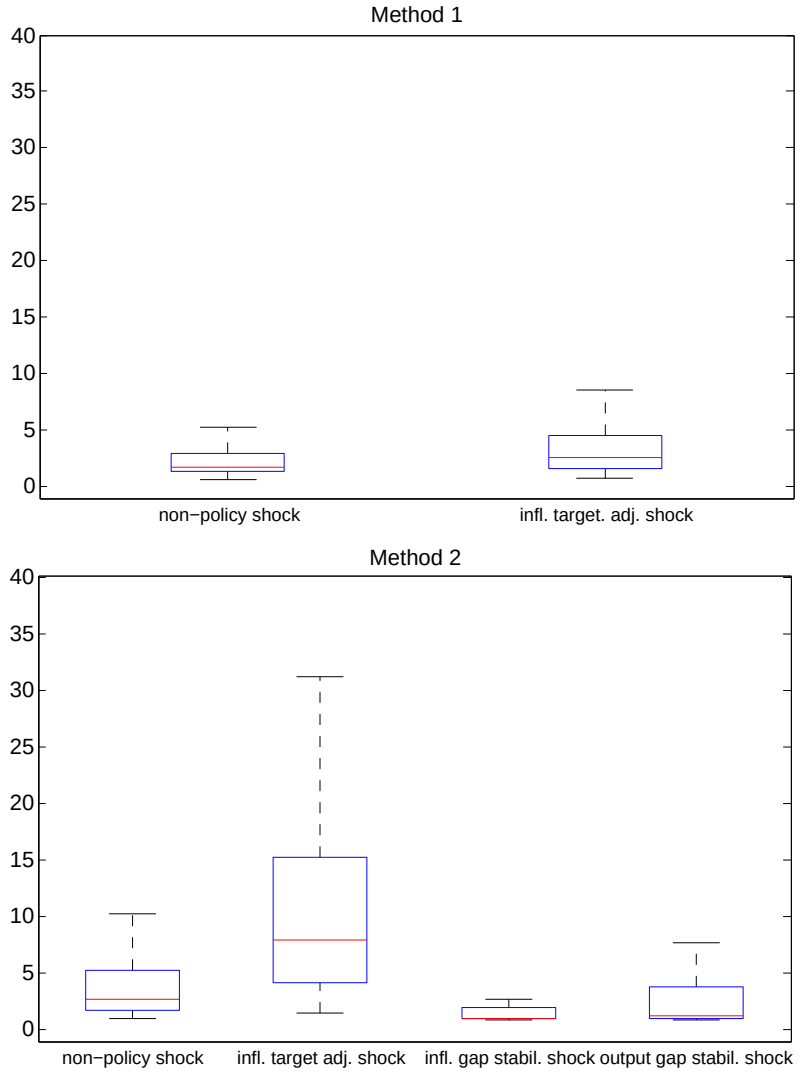
$$1 + 2\sum_{j=1}^{J} \rho_j,$$

where ρ_j is the sample autocorrelation at lag j through lag J corresponding to all MCMC draws for a specific parameter. In our empirical application we set J to be large enough until autocorrelation tapers off. To give an idea on how to assess the metric above, one should recall that if MCMC draws were indeed independent draws (i.e., perfect mixing of the chain), then the inefficiency factor would be 1. As a rule of thumb, an inefficiency factor of 100 means that approximately 10000 posterior draws are required to convey the same information as 100 independent draws. Table 3.7 reports the inefficiency factors for parameter draws. Figure 3.7 report inefficiency factors for the $T \times 1$ innovations associated with policy and non-policy shocks discussed in Section 3.3. For innovations, instead of reporting inefficiency factors for each one of the T elements in $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$, we use boxplots to present the information on inefficiency factor visually. In particular, the middle line denotes the median inefficiency factor based on a sample T inefficiency factors constructed for each innovation at point t. Lower and upper lines represent the 25 and 75 percentiles, respectively. Whiskers extend to the maximum and minimum inefficiency factor computed from all draws. All and in all, our results below suggest that the MCMC algorithm developed in this chapter has good mixing properties. For example, results for method 2, which entails regime switching, are broadly comparable with values reported in Fiorentini et al. [2014] who use a variant of the mixture innovation of Gerlach et al. [2000] to sample discrete states.

Table 3.7: Inefficiency factors for the parameters sampled using the MCMC algorithms in Section 3.4.2.1. Results for methods 1 correspond to model M_{R1}.

Parameter	Description	Method 1 (M_{R1})	Method 2
τ_0	Initial condition for trend inflation	135.19	170.09
$\beta_{0,0}$	Initial condition for drifting intercept in the policy rule	139.04	168.03
$\beta_{1,0}$	Initial condition for drifting coefficient for inflation expectation	–	82.36
$\beta_{2,0}$	Initial condition for drifting coefficient for output gap expectation	–	129.15
β_1	Inflation expectation coefficient	38.81	–
β_2	Output gap expectation coefficient	52.03	–
β_3	Money growth expectation coefficient	12.23	40.11
β_4	Policy smoothing coefficient in the policy rule	47.86	112.06
γ_1	Transmission of a systematic policy shock to τ_t during the Great Inflation (method 1)	98.15	–
γ_2	Transmission of a systematic policy shock to τ_t during the Great Moderation (method 1)	53.15	–
$\gamma_1^{\pi^*}$	Transmission of a systematic policy shock (π_t^*) to τ_t during the Great Inflation (method 2)	–	116.46
$\gamma_2^{\pi^*}$	Transmission of a systematic policy shock (π_t^*) to τ_t during the Great Moderation (method 2)	–	42.85
$\gamma_1^{\omega_1}$	Transmission of a systematic policy shock ($\omega_{1,t}$) to τ_t during the Great Inflation (method 2)	–	3.72
$\gamma_2^{\omega_1}$	Transmission of a systematic policy shock ($\omega_{1,t}$) to τ_t during the Great Moderation (method 2)	–	2.62
$\gamma_1^{\omega_2}$	Transmission of a systematic policy shock ($\omega_{2,t}$) to τ_t during the Great Inflation (method 2)	–	1.76
$\gamma_2^{\omega_2}$	Transmission of a systematic policy shock ($\omega_{2,t}$) to τ_t during the Great Moderation (method 2)	–	1.18
$\sigma^2_{\pi,1}$		15.57	26.77
$\sigma^2_{r,1}$	Second moment parameters in $\Omega_{e,1}$ during the Great Inflation	10.12	130.04
$\sigma_{\pi r,1}$		20.03	65.40
$\sigma^2_{\pi,2}$		16.89	8.10
$\sigma^2_{r,2}$	Second moment parameters in $\Omega_{e,2}$ during the Great Moderation	7.17	11.49
$\sigma_{\pi r,2}$		12.85	22.90
$\sigma^2_{v_{\tau^*},1}$	Variance of non-policy shock to trend inflation during the Great Inflation	42.67	46.03
$\sigma^2_{v_{\pi^*},1}$	Variance of inflation target adjustment shock during the Great Inflation	3.56	4.01
$\sigma^2_{v_{\omega_1},1}$	Variance of inflation gap stabilization shock during the Great Inflation	–	6.73
$\sigma^2_{v_{\omega_2},1}$	Variance of output gap stabilization shock during the Great Inflation	–	5.53
$\sigma^2_{v_{\tau^*},2}$	Variance of non-policy shock to trend inflation during the Great Moderation	2.11	4.15
$\sigma^2_{v_{\pi^*},2}$	Variance of inflation target adjustment shock during the Great Moderation	4.00	13.17
$\sigma^2_{v_{\omega_1},2}$	Variance of inflation gap stabilization shock during the Great Moderation	–	3.25
$\sigma^2_{v_{\omega_2},2}$	Variance of output gap stabilization shock during the Great Moderation	–	3.16
$p_{1,1}$		–	32.47
$p_{1,2}$		–	32.39
$p_{1,3}$		–	38.13
$p_{2,1}$		–	37.06
$p_{2,2}$	Transition probability, $p_{i,j}$, between identification strategy i and j, for $i, j = 1, 2, 3$.	–	16.31
$p_{2,3}$		–	21.65
$p_{3,1}$		–	43.32
$p_{3,2}$		–	47.13
$p_{3,3}$		–	21.82

Figure 3.7: Inefficiency factors for the innovations sampled using the MCMC algorithms in Section 3.4.2.1. Results for method 1 correspond to model M_{R1}



Chapter 4

Conclusion

4.1 Main findings

The focus of this thesis has been on developing methodological contributions for Bayesian inference of state space models. In particular, we investigated how relaxing the full rank assumption for the covariance matrix of the errors driving the states can be useful for macroeconometric modeling. The techniques developed in this thesis were illustrated in the context of inflation modeling applications.

Chapter 2 explores the idea of rank reduction of the covariance matrix within unobserved components models. Point and density forecasting implications that arise when some state variables underpinning a time series decomposition are driven by the same disturbance (i.e., perfect correlation) are investigated within a Bayesian framework. The analysis in Chapter 2 focuses on univariate unobserved components models. In addition to the comprehensive discussion of the relationship between state correlation and Bayesian forecasting, three other contributions are presented in Chapter 2: (1) a new class of unobserved components models that bridges the existing orthogonal and perfectly correlated innovations schemes; (2) a new algorithm for fast estimation of state space models based on precision sampling techniques; and (3) a substantial inflation forecasting application incorporating various specifications of unobserved components models.

Chapter 3 extends the idea of rank reduction of the covariance matrix to a multivariate framework. In particular, we explore how reducing the number of sources driving a particular state space representation can be used to assess the contribution of changes in the conduct of monetary policy for U.S. trend inflation. Overall, our results in Chapter 3 suggest that the increased influence monetary policy exerted over trend inflation during the Great Moderation reflects a combination of monetary policy reacting more aggressively to inflation gap movements and less aggressively to output gap movements towards the end of the Great Inflation period. In addition to the empirical contribution, Chapter 3 also develops a Markov Chain Monte Carlo

algorithm for state space models with regime switching. In particular, the switching mechanism studied in this chapter explores different state correlation structures. Including correlation structures which lead to models with full and reduced rank covariance matrices. Consequently, the modeling framework in Chapter 3 provides a flexible environment which enables endogenous determination of the covariance matrix rank.

Bibliography

Anderson, T. W., 1984. *An Introduction to Multivariate Statistical Analysis*. Wiley. (cited on page 41)

Ascari, G. and Sbordone, A. M., 2014. The Macroeconomics of Trend Inflation. *Journal of Economic Literature*, 52, 3 (2014), 679–739. (cited on pages 54 and 75)

Bai, J. and Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, (1998), 47–78. (cited on page 73)

Bauwens, L.; Koop, G.; Korobilis, D.; and Rombouts, J. V., 2014. The Contribution of Structural Break Models to Forecasting Macroeconomic Series. *Journal of Applied Econometrics*, (2014). (cited on pages 6 and 27)

Berg, A.; Meyer, R.; and Yu, J., 2004. Deviance Information Criterion for Comparing Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 22, 1 (2004), 107–120. (cited on page 70)

Bernanke, B. S.; Gertler, M.; Watson, M.; Sims, C. A.; and Friedman, B. M., 1997. Systematic Monetary Policy and the Effects of Oil Price Shocks. *Brookings papers on economic activity*, 1997, 1 (1997), 91–157. (cited on page 60)

Bernanke, B. S. and Mihov, I., 1998. The Liquidity Effect and Long-Run Neutrality. In *Carnegie-Rochester conference series on public policy*, vol. 49, 149–194. Elsevier. (cited on page 60)

Beveridge, S. and Nelson, C. R., 1981. A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle. *Journal of Monetary Economics*, 7, 2 (1981), 151–174. (cited on pages 4, 7, and 53)

Binder, C. C., 2015. Fed Speak on Main Street. *UC Berkeley, Manuscript*, (2015). (cited on page 54)

Boivin, J., 2006. Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data. *Journal of Money, Credit and Banking*, 38, 5 (2006), 1149–1173. (cited on pages 55, 58, 61, 62, 80, 87, and 92)

Boivin, J. and Giannoni, M. P., 2006. Has Monetary Policy Become More Effective? *The Review of Economics and Statistics*, 88, 3 (2006), 445–462. (cited on pages 60 and 61)

Boldea, O. and Hall, A. R., 2013. Estimation and Inference in Unstable Nonlinear Least Squares Models. *Journal of Econometrics*, 172, 1 (2013), 158–167. (cited on page 73)

Canova, F. and Ferroni, F., 2012. The Dynamics of US Inflation: Can Monetary Policy Explain the Changes? *Journal of Econometrics*, 167, 1 (2012), 47–60. (cited on pages 58 and 84)

Carlin, B. P. and Louis, T. A., 1997. Bayes and Empirical Bayes Methods for Data Analysis. *Statistics and Computing*, 7, 2 (1997), 153–154. (cited on page 70)

Carter, C. K. and Kohn, R., 1994. On Gibbs Sampling for State Space Models. *Biometrika*, 81, 3 (1994), 541–553. (cited on pages 16 and 76)

Casals, J.; Garcia-Hiernaux, A.; and Jerez, M., 2012. From General State-Space to VARMAX Models. *Mathematics and Computers in Simulation*, 82, 5 (2012), 924–936. (cited on page 73)

Casals, J. and Sotoca, S., 2001. The Exact Likelihood for a State Space Model with Stochastic Inputs. *Computers & Mathematics with Applications*, 42, 1 (2001), 199–209. (cited on page 18)

Casals, J.; Sotoca, S.; and Jerez, M., 2015. Single and multiple error state-space models for signal extraction. *Journal of Statistical Computation and Simulation*, 85, 5 (2015), 1053–1069. (cited on pages 16 and 77)

Castelnuovo, E., 2010. Trend Inflation and Macroeconomic Volatilities in the Post-WWII US Economy. *The North American Journal of Economics and Finance*, 21, 1 (2010), 19–33. (cited on page 54)

Chan, J. C., 2013. Moving Average Stochastic Volatility Models with Application to Inflation Forecast. *Journal of Econometrics*, 176, 2 (2013), 162–172. (cited on pages 4, 5, 8, 17, 25, 30, 53, and 97)

Chan, J. C., 2015. The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling. *Journal of Business & Economic Statistics*, , forthcoming (2015). (cited on pages 5 and 37)

Chan, J. C.; Clark, T. E.; and Koop, G., 2015. A New Model of Inflation, Trend Inflation, and Long-Run Inflation Expectations. *Federal Reserve System of Cleveland, Working Paper*, 5 (2015). (cited on page 53)

Chan, J. C. and Grant, A. L., 2014. Fast Computation of the Deviance Information Criterion for Latent Variable Models. *Computational Statistics & Data Analysis*, (2014). (cited on page 107)

Chan, J. C. and Jeliazkov, I., 2009. Efficient Simulation and Integrated Likelihood Estimation in State Space Models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 1-2 (2009), 101–120. (cited on pages 6, 17, 21, 56, 66, 77, and 97)

Chan, J. C.; Koop, G.; and Potter, S. M., 2013. A New Model of Trend Inflation. *Journal of Business & Economic Statistics*, 31, 1 (2013), 94–106. (cited on pages 53 and 55)

Chatfield, C.; Koehler, A. B.; Ord, J. K.; and Snyder, R. D., 2001. A New Look at Models for Exponential Smoothing. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50, 2 (2001), 147–159. (cited on pages 5 and 8)

Chib, S., 1996. Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models. *Journal of Econometrics*, 75, 1 (1996), 79–7. (cited on pages 75, 79, and 103)

Chib, S., 2001. Markov Chain Monte Carlo Methods: Computation and Inference. *Handbook of Econometrics*, 5 (2001), 3569–3649. (cited on pages 35 and 114)

Christiano, L. J.; Eichenbaum, M.; and Evans, C. L., 2005. Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, 113, 1 (2005), 1–45. (cited on page 8)

Clarida, R.; Galí, J.; and Gertler, M., 2000. Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics*, 115, 1 (2000), 147–180. (cited on pages 60 and 61)

Clark, P. K., 1987. The Cyclical Component of US Economic Activity. *The Quarterly Journal of Economics*, (1987), 797–814. (cited on pages 3 and 10)

Clark, T. E. and Doh, T., 2014. Evaluating Alternative Models of Trend Inflation. *International Journal of Forecasting*, 30, 3 (2014), 426–448. (cited on pages 5, 6, 8, 25, 27, 30, 53, and 57)

Clark, T. E. and Ravazzolo, F., 2014. Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, (2014). (cited on pages 6 and 27)

Cochrane, J. H., 1988. How Big is the Random Walk in GNP? *The Journal of Political Economy*, (1988), 893–920. (cited on page 13)

Cogley, T.; Morozov, S.; and Sargent, T. J., 2005. Bayesian Fan Charts for UK Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System. *Journal of Economic Dynamics and Control*, 29, 11 (2005), 1893–1925. (cited on page 27)

Cogley, T.; Primiceri, G. E.; and Sargent, T. J., 2010. Inflation-Gap Persistence in the US. *American Economic Journal: Macroeconomics*, 2, 1 (2010), 43–69. (cited on pages 54 and 58)

Cogley, T. and Sargent, T. J., 2001. Evolving Post-World War II US Inflation Dynamics. *NBER Macroeconomics Annual 2001, Volume 16*, (2001), 331–388. (cited on pages 54, 58, and 74)

Cogley, T. and Sargent, T. J., 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US. *Review of Economic dynamics*, 8, 2 (2005), 262–302. (cited on pages 54, 55, 58, 61, 62, and 74)

Cogley, T. and Sbordone, A. M., 2008. Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve. *The American Economic Review*, 98, 5 (2008), 2101–2126. (cited on page 54)

Coibion, O. and Gorodnichenko, Y., 2011. Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation. *American Economic Review*, 101, 1 (February 2011), 341–70. (cited on pages 54, 55, 62, 84, and 87)

De Jong, P. and Chu-Chun-Lin, S., 1994. Fast Likelihood Evaluation and Prediction for Nonstationary State Space Models. *Biometrika*, 81, 1 (1994), 133–142. (cited on page 18)

De Jong, P. and Shephard, N., 1995. The Simulation Smoother for Time Series Models. *Biometrika*, 82, 2 (1995), 339–350. (cited on page 16)

Diebold, F. X. and Mariano, R. S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, (1995). (cited on pages 27, 31, and 33)

Dungey, M.; Jacobs, J. P.; Tian, J.; and van Norden, S., 2015. Trend in Cycle or Cycle in Trend? New Structural Identifications for Unobserved-Components Models of U.S. Real GDP. *Macroeconomic Dynamics*, 19 (6 2015), 776–790. (cited on page 4)

Durbin, J. and Koopman, S. J., 2002. A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika*, 89, 3 (2002), 603–616. (cited on page 16)

Durbin, J. and Koopman, S. J., 2012. *Time Series Analysis by State Space Methods*. 38. Oxford University Press. (cited on pages 17, 66, 74, and 76)

Eisenstat, E. and Strachan, R. W., 2014. Modelling Inflation Volatility. Technical report, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University. (cited on page 70)

Fiorentini, G.; Planas, C.; and Rossi, A., 2014. Efficient MCMC sampling in Dynamic Mixture Models. *Statistics and Computing*, 24, 1 (2014), 77–89. (cited on pages 76 and 114)

Forbes, C. S.; Snyder, R. D.; and Shami, R. G., 2000. Bayesian Exponential Smoothing. *Unpublished Manuscript, Monash University, Department of Econometrics and Business Statistics*, (2000). (cited on page 14)

Frühwirth-Schnatter, S., 1994. Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis*, 15, 2 (1994), 183–202. (cited on page 16)

Frühwirth-Schnatter, S., 2001. Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association*, 96, 453 (2001), 194–209. (cited on page 73)

Frühwirth-Schnatter, S., 2004. Efficient Bayesian Parameter Estimation for State Space Models Based on Reparameterizations. *State Space and Unobserved Component Models: Theory and Applications*, (2004), 123–151. (cited on page 35)

Frühwirth-Schnatter, S. and Wagner, H., 2010. Stochastic Model Specification Search for Gaussian and Partial Non-Gaussian State Space Models. *Journal of Econometrics*, 154, 1 (2010), 85–100. (cited on pages 10 and 35)

Gagnon, J.; Raskin, M.; Remache, J.; and Sack, B. P., 2010. Large-Scale Asset Purchases by the Federal Reserve: Did They Work? *FRB of New York Staff Report*, , 441 (2010). (cited on page 55)

Galí, J. and Gambetti, L., 2009. On the Sources of the Great Moderation. *American Economic Journal: Macroeconomics*, 1, 1 (2009), 26–57. (cited on pages 54 and 84)

Gamerman, D. and Lopes, H. F., 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press. (cited on page 14)

Garnier, C.; Mertens, E.; and Nelson, E., 2015. Trend Inflation in Advanced Economies. *International Journal of Central Banking*, 11, S1 (2015), 65–136. (cited on pages 5, 6, 7, 8, 25, 27, 53, 54, 55, and 58)

Gelfand, A. E. and Dey, D. K., 1994. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, (1994), 501–514. (cited on page 109)

Gerlach, R.; Carter, C.; and Kohn, R., 2000. Efficient Bayesian Inference for Dynamic Mixture Models. *Journal of the American Statistical Association*, 95, 451 (2000), 819–828. (cited on pages 76, 77, and 114)

Geweke, J., 1992. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In *Bernardo, J., Berger, J., Dawid, A. and Smith, A. (eds.), Bayesian Statistics 4*, 641–649. Oxford: Clarendon Press. (cited on page 27)

Geweke, J., 1999. Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication. *Econometric reviews*, 18, 1 (1999), 1–73. (cited on pages 83 and 109)

Geweke, J., 2007. Interpretation and Inference in Mixture Models: Simple MCMC Works. *Computational Statistics & Data Analysis*, 51, 7 (2007), 3529–3550. (cited on page 73)

Geweke, J. and Amisano, G., 2011. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26, 1 (2011), 1–29. (cited on page 32)

Geweke, J. and Whiteman, C., 2006. Bayesian forecasting. *Handbook of Economic Forecasting*, 1 (2006), 3–80. (cited on page 12)

Giordani, P. and Kohn, R., 2008. Efficient Bayesian inference for Multiple Change-Point and Mixture Innovation Models. *Journal of Business & Economic Statistics*, 26, 1 (2008). (cited on page 77)

Golub, G. H. and Van Loan, C. F., 1983. *Matrix Computations*, vol. 3. The Johns Hopkins University Press, Baltimore. (cited on pages 21 and 22)

Goutis, C. and Casella, G., 1999. Explaining the Saddlepoint Approximation. *The American Statistician*, 53, 3 (1999), 216–224. (cited on page 13)

Granger, C. W. J. and Newbold, P., 1986. *Forecasting Economic Time Series*. Academic Press. (cited on pages 12, 40, and 73)

Hamilton, J. D.; Waggoner, D. F.; and Zha, T., 2007. Normalization in econometrics. *Econometric Reviews*, 26, 2-4 (2007), 221–252. (cited on page 73)

Harvey, A. and Koopman, S. J., 2000. Signal Extraction and the Formulation of Unobserved Components Models. *The Econometrics Journal*, 3, 1 (2000), 84–107. (cited on pages 3, 4, 5, and 16)

Harvey, A. C., 1985. Trends and Cycles in Macroeconomic Time Series. *Journal of Business & Economic Statistics*, 3, 3 (1985), 216–227. (cited on pages 3 and 5)

Harvey, A. C., 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. (cited on pages 8, 17, 42, 66, and 74)

Harvey, A. C. and Jaeger, A., 1993. Detrending, Stylized Facts and the Business Cycle. *Journal of Applied Econometrics*, 8 (1993), 231–231. (cited on pages 10 and 25)

Ireland, P. N., 2004. Money's Role in the Monetary Business Cycle. Technical Report 6. (cited on page 58)

Ireland, P. N., 2007. Changes in the Federal Reserve's Inflation Target: Causes and Consequences. *Journal of Money, Credit and Banking*, 39, 8 (2007), 1851–1882. (cited on pages 54, 75, and 84)

Iwata, S. and Li, H., 2015. What are the Differences in Trend Cycle Decompositions by Beveridge and Nelson and by Unobserved Component Models? *Econometric Reviews*, 34, 1-2 (2015), 146–173. (cited on page 4)

Kang, K. H.; Kim, C.-J.; and Morley, J., 2009. Changes in US Inflation Persistence. *Studies in Nonlinear Dynamics & Econometrics*, 13, 4 (2009). (cited on pages 7, 8, and 55)

Kass, R. E. and Raftery, A. E., 1995. Bayes Factors. *Journal of the American Statistical Association*, 90, 430 (1995), 773–795. (cited on pages xv and 82)

Kiley, M. T., 2014. The Aggregate Demand Effects of Short-and Long-Term Interest Rates. *International Journal of Central Banking*, 10, 4 (2014), 69–104. (cited on page 55)

Kim, C.-J. and Kim, J., 2013a. The Pile-up Problem in Trend-Cycle Decomposition of Real GDP: Classical and Bayesian Perspectives. (2013). (cited on page 73)

Kim, C.-J. and Kim, J., 2013b. The âĂŸPile-up Problem'in Trend-Cycle Decomposition of Real GDP: Classical and Bayesian Perspectives. (2013). (cited on page 74)

Kim, C.-J. and Nelson, C. R., 1999. Has the US Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle. *Review of Economics and Statistics*, 81, 4 (1999), 608–616. (cited on pages 54, 58, and 75)

Kim, C.-J. and Nelson, C. R., 2006. Estimation of a Forward-Looking Monetary Policy Rule: A Time-Varying Parameter Model Using Ex Post Data. *Journal of Monetary Economics*, 53, 8 (2006), 1949–1966. (cited on pages 55, 58, 62, and 87)

Kim, S.; Shephard, N.; and Chib, S., 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65, 3 (1998), 361–393. (cited on page 4)

Koop, G., 2003. *Bayesian Econometrics*. Wiley. (cited on pages 14, 23, and 82)

Koop, G. and Korobilis, D., 2010. *Bayesian Multivariate Time Series Methods for Empirical Macroeconomics*. Now Publishers Inc. (cited on page 75)

Koop, G.; Leon-Gonzalez, R.; and Strachan, R. W., 2009. On the Evolution of the Monetary Policy Transmission Mechanism. *Journal of Economic Dynamics and Control*, 33, 4 (2009), 997–1017. (cited on pages 74 and 75)

Koop, G.; Poirier, D. J.; and Tobias, J. L., 2007. *Bayesian Econometric Methods*, vol. 7. Cambridge University Press. (cited on pages 21, 77, 99, and 100)

Kumar, S.; Afrouzi, H.; Coibion, O.; and Gorodnichenko, Y., 2015. Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand. (2015). (cited on page 54)

Laubach, T. and Williams, J. C., 2003. Measuring the Natural Rate of Interest. *Review of Economics and Statistics*, 85, 4 (2003), 1063–1070. (cited on pages 61 and 74)

Leeper, E. M. and Leith, C., 2016. Understanding Inflation as a Joint Monetary-Fiscal Phenomenon. *National Bureau of Economic Research, working paper*, (2016). (cited on page 55)

Levin, A. T., 2014. The Design and Communication of Systematic Monetary Policy Strategies. *Journal of Economic Dynamics and Control*, 49 (2014), 52–69. (cited on page 60)

Luo, S. and Startz, R., 2014. Is it One Break or Ongoing Permanent Shocks that Explains US Real GDP? *Journal of Monetary Economics*, 66 (2014), 155–163. (cited on pages 3, 9, and 70)

Marcellino, M.; Stock, J. H.; and Watson, M. W., 2006. A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics*, 135, 1 (2006), 499–526. (cited on page 25)

McCausland, W. J.; Miller, S.; and Pelletier, D., 2011. Simulation Smoothing For State-Space Models: A Computational Efficiency Analysis. *Computational Statistics & Data Analysis*, 55, 1 (2011), 199–212. (cited on pages 6 and 66)

McConnell, M. M. and Perez-Quiros, G., 2000. Output Fluctuations in the United States: What Has Changed Since the Early 1980's? *American Economic Review*, 90, 5 (2000), 1464–1476. (cited on pages 58 and 75)

Meyer, L. H. and Bomfim, A. N., 2010. Quantifying the Effects of Fed Asset Purchases on Treasury Yields. *Monetary Policy Insights: Fixed Income Focus*, (2010). (cited on page 55)

Morley, J., 2015. Discussion of âĂIJTrend Inflation in Advanced EconomiesâĂİ. *International Journal of Central Banking*, 11, S1 (2015), 137–143. (cited on page 54)

Morley, J. and Piger, J., 2012. The Asymmetric Business Cycle. *Review of Economics and Statistics*, 94, 1 (2012), 208–221. (cited on page 70)

Morley, J. C.; Nelson, C. R.; and Zivot, E., 2003. Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP so Different? *Review of Economics and Statistics*, 85, 2 (2003), 235–243. (cited on pages 3, 4, 5, and 8)

Oh, K. H. and Zivot, E., 2006. The Clark Model with Correlated Components. *Unpublished Manuscript, Department of economics, University of Washington*, (2006). (cited on pages 4 and 10)

Oh, K. H.; Zivot, E.; and Creal, D., 2008. The Relationship Between the Beveridge-Nelson Decomposition and Other Permanent-Transitory Decompositions that Are Popular in Economics. *Journal of Econometrics*, 146, 2 (2008), 207–219. (cited on pages 4 and 5)

Ord, J. K.; Koehler, A.; and Snyder, R. D., 1997. Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, 92, 440 (1997), 1621–1629. (cited on pages 5 and 8)

Ord, J. K.; Snyder, R. D.; Koehler, A. B.; Hyndman, R. J.; and Leeds, M., 2005. Time Series Forecasting: The Case for the Single Source of Error State Space Approach. *Unpublished Manuscript, Working Paper 7/05, Monash Econometrics and Business Statistics*, (2005). (cited on page 5)

Orphanides, A., 2001. Monetary Policy Rules Based on Real-Time Data. *American Economic Review*, (2001), 964–985. (cited on pages 58, 60, and 87)

Orphanides, A., 2004. Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches. *Journal of Money, credit, and Banking*, 36, 2 (2004), 151–175. (cited on pages 58, 60, 61, and 80)

Papaspiliopoulos, O.; Roberts, G.; and Söld, M., 2003. Non-Centered Parameterisations for Hierarchical Models and Data Augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, 307–326. Oxford University Press, USA. (cited on page 35)

Perron, P. and Wada, T., 2009. Let's Take a Break: Trends and Cycles in US Real GDP. *Journal of Monetary Economics*, 56, 6 (2009), 749–765. (cited on page 3)

Piger, J. and Rasche, R., 2008. Inflation: Do Expectations Trump the Gap? *International Journal ofCentral Banking*, 4 (2008), 851–16. (cited on page 58)

Pollock, D. S. G.; Green, R. C.; and Nguyen, T., 1999. *Handbook of Time Series Analysis, Signal Processing, and Dynamics*. Academic Press. (cited on pages 18 and 78)

Primiceri, G. E., 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72, 3 (2005), 821–852. (cited on pages 55, 58, 61, 62, 74, and 75)

Proietti, T., 2006. Trend-cycle Decompositions with Correlated Components. *Econometric Reviews*, 25, 1 (2006), 61–84. (cited on pages 3 and 4)

Raftery, A. E., 1995. Bayesian Model Selection in Social Research. *Sociological methodology*, (1995), 111–163. (cited on pages xv and 82)

Rudebusch, G. D., 2002. Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia. *Journal of Monetary Economics*, 49, 6 (2002), 1161–1187. (cited on page 61)

Sargan, J. D. and Bhargava, A., 1983. Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk. *Econometrica: Journal of the Econometric Society*, (1983), 153–174. (cited on page 73)

Sims, C. A. and Zha, T., 1998. Bayesian methods for dynamic multivariate models. *International Economic Review*, (1998), 949–968. (cited on page 74)

Sims, C. A. and Zha, T., 2006. Were There Regime Switches in US Monetary Policy? *The American Economic Review*, 96, 1 (2006), 54–81. (cited on pages 60 and 75)

Smets, F. and Wouters, R., 2007. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97, 3 (2007), 586–606. (cited on pages 8 and 84)

Snyder, R., 1985. Recursive Estimation of Dynamic Linear Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, (1985), 272–276. (cited on page 5)

Snyder, R. D.; Ord, J. K.; and Koehler, A. B., 2001. Prediction Intervals for ARIMA Models. *Journal of Business & Economic Statistics*, 19, 2 (2001), 217–225. (cited on pages 8, 14, and 18)

Spiegelhalter, D. J.; Best, N. G.; Carlin, B. P.; and Van Der Linde, A., 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 4 (2002), 583–639. (cited on page 56)

Stock, J. H. and Watson, M. W., 1998. Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model. *Journal of the American Statistical Association*, 93, 441 (1998), 349–358. (cited on pages 73 and 74)

Stock, J. H. and Watson, M. W., 2003. Has the Business Cycle Changed and Why? In *NBER Macroeconomics Annual 2002, Volume 17*, 159–230. MIT press. (cited on pages 54, 58, and 75)

Stock, J. H. and Watson, M. W., 2007. Why Has US Inflation Become Harder to Forecast? *Journal of Money, Credit and banking*, 39, s1 (2007), 3–33. (cited on pages 3, 4, 5, 8, 25, 39, 53, 55, 58, and 93)

Stock, J. H. and Watson, M. W., 2015. Core Inflation and Trend Inflation. *Review of Economics and Statistics*, , 0 (2015). (cited on page 53)

Taylor, J. B., 1993. Discretion versus Policy Rules in Practice. *Carnegie-Rochester conference series on public policy*, 39 (1993), 195–214. (cited on pages 55 and 74)

Teräsvirta, T., 1977. The Invertibility of Sums of Discrete MA and ARMA Processes. *Scandinavian Journal of Statistics*, (1977), 165–170. (cited on page 40)

Tierney, L., 1994. Markov Chains for Exploring Posterior Distributions. *the Annals of Statistics*, (1994), 1701–1728. (cited on page 24)

Verdinelli, I. and Wasserman, L., 1995. Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, 90, 430 (1995), 614–618. (cited on page 81)

Watson, M. W., 1986. Univariate Detrending Methods with Stochastic Trends. *Journal of Monetary Economics*, 18, 1 (1986), 49–75. (cited on pages 3 and 42)

Winters, P. R., 1960. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, 6, 3 (1960), 324–342. (cited on page 18)

Woodford, M., 2007. How Important is Money in the Conduct of Monetary Policy? *Journal of Money, Credit and Banking*, 40, 8 (2007), 1561–1598. (cited on pages 10 and 54)

Zarnowitz, V. and Ozyildirim, A., 2006. Time Series Decomposition and Measurement of Business Cycles, Trends and Growth Cycles. *Journal of Monetary Economics*, 53, 7 (2006), 1717–1739. (cited on pages 10 and 25)