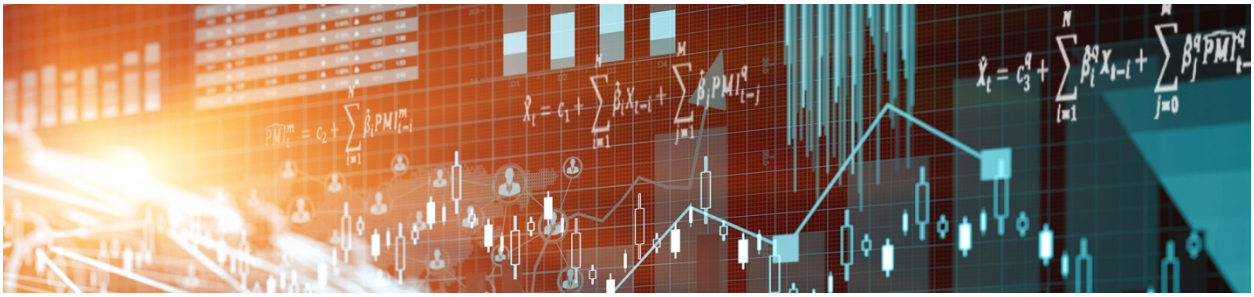
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2018-14

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



by Luis Uzeda

Bank of Canada Staff Working Paper 2018-14

March 2018

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

by

Luis Uzeda

Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
LUzedaGarcia@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I would like to thank Joshua Chan, Catherine Forbes, Alfred Haug, Jan Jacobs, James Morley, Rodney Strachan and Tomasz Wozniak for helpful comments and suggestions. I also thank seminar and conference participants at the 2016 Australasia Meeting of the Econometric Society, the 2nd Workshop of the Australasian Macroeconomic Society, the 2nd Continuing Education in Macroeconometrics Workshop at the University of Tasmania, the Australian National University and the 28th PhD in Economics and Business Conference at the University of Queensland. The views expressed in this paper are mine and do not necessarily reflect the position of the Bank of Canada.

Abstract

Implications for signal extraction from specifying unobserved components (UC) models with correlated or orthogonal innovations have been well investigated. In contrast, the forecasting implications of specifying UC models with different state correlation structures 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, and a new approach that 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.

Bank topics: Econometric and statistical methods; Inflation and prices
JEL codes: C11, C15, C51, C53

Résumé

Au regard de l'extraction des signaux, les implications de la spécification de modèles à composantes non observées avec innovations corrélées ou orthogonales ont été largement analysées. Par contraste, s'agissant des prévisions, les implications de modèles à composantes non observées avec différentes structures de corrélation des variables d'état sont moins bien comprises. Dans le présent article, nous tentons de combler cette lacune, à la lumière de la récente résurgence d'études qui adoptent des modèles à composantes non observées afin d'établir des prévisions. Nous considérons quatre structures de corrélation entre les innovations : orthogonales, corrélées et parfaitement corrélées ainsi qu'une nouvelle approche qui combine des éléments de deux structures très différentes, à savoir les innovations orthogonales et parfaitement corrélées. Nous analysons, dans un cadre bayésien, les restrictions de l'espace de paramètres associées à différentes structures de corrélation et leurs liens avec les prévisions. Comme les innovations parfaitement corrélées réduisent le rang de la matrice de covariance, nous développons un algorithme d'échantillonnage de Monte-Carlo par chaînes de Markov qui se fonde sur les propriétés des matrices de Toeplitz et les récentes avancées dans le domaine des algorithmes de précision. Selon nos résultats de plusieurs mesures de l'inflation aux États-Unis, la structure de corrélation entre les variables d'état a d'importantes implications pour la qualité des prévisions et les estimations de l'inflation tendancielle.

Sujets : Méthodes économétriques et statistiques; Inflation et prix
Codes JEL : C11, C15, C51, C53

Non-technical summary

Unobserved components (UC) models have been widely employed in empirical macroeconomic studies over the years. This can be largely attributed to the fact that UC models provide an appropriate framework to capture features of economic time series that are instrumental for researchers and policy makers. For example, time series properties – such as the existence of trends and cyclical and seasonal patterns – can be naturally accommodated within a UC framework and extracted using filtering techniques and related methods.

Over the past three decades, a wealth of studies analyzed the in-sample properties of UC models. In particular, the issue of specifying the correlation between trend and cyclical dynamics in a UC framework proved to be important for business cycle measurement (Morley, Nelson and Zivot (2003), Oh, Zivot and Creal (2008) and Iwata and Li (2015)) and for investigating properties of parametric filters (Harvey and Koopman (2000) and Proietti (2006)). However, the out-of-sample implications of specifying different correlation structures between permanent and transitory movements remain less well understood. In view of the recent resurgence of studies adopting UC models for forecasting purposes, this paper attempts to address this gap.

Overall, three main contributions arise from this study: (1) a substantial forecasting exercise for several measures of U.S. inflation based on eleven classes of UC models; (2) a new class of UC model that accommodates two contrasting correlation structures (absence of and perfect correlation); and (3) an efficient estimation algorithm that nests all UC models adopted in this paper. Our results indicate that allowing for correlation between permanent and transitory price level movements lead to some improvements in inflation forecasting at both point and density dimensions. We also document that measures of trend inflation can be sensitive to how one specifies the correlation between permanent and transitory price level movements.

1 Introduction

Unobserved components (UC) models provide a flexible and yet parsimonious framework that 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 of a signal extraction problem). For example, Harvey and 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 between business cycle measures generated by UC models and measures 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 U.S. 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 paper contributes to fill this gap. Moreover, the recent resurgence of papers adopting UC models for inflation forecasting purposes – as in Stock and Watson (2007), Stella and Stock (2013), Chan (2013), Chan et al. (2013), Clark and Doh (2014), Garnier et al. (2015) and Stock and Watson (2016) – strengthens the case for out-of-sample evaluation of the correlation structure within such models.

Notably, modern approaches to forecasting 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) using orthogonal innovations; and (3) introducing stochastic volatility à la, e.g., Kim et al. (1998) to model changes in the conditional variance of innovations over time. In this paper we keep point 1, extend point 2 and leave point 3 for future research. To be clear, in our empirical exercise, we leave out stochastic volatility not because we

[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.

[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.

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 that 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 that incorporate parameter uncertainty, as in the case of Bayesian estimation, the correlation structure has a direct connection with the forecasting function that 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 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 with 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 that we address in this paper.

Our empirical evaluation is built around four correlation structures: innovations (or, equivalently, states) are allowed to be orthogonal, correlated, or perfectly correlated, but we also include a new approach combining aspects from two contrasting correlation structures. In particular, we construct UC models that 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 that 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 model as *reduced source of error* (RSOE) to distinguish it from its SSOE and *multiple source of error* (MSOE) counterparts. Since parameters in both MSOE with orthogonal innovations and SSOE models are identified, RSOE variants provide a novel strategy to allow for state correlation without incurring well-known limitations of identifying covariance matrix parameters within UC models (see, e.g., Morley et al. (2003), Oh and Zivot (2006) and Oh et al. (2008)) and without requiring perfect correlation across all state variables. 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. As we show later, the RSOE scheme can also represent a compromise between SSOE and MSOE variants in terms of

parameter 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 for 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 that builds upon properties of Toeplitz matrices and extends 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 that adds to the existing (Kalman-filter based) ones of De Jong 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 paper is as follows: Section 2 presents all UC models entertained in this paper. In Section 3 we discuss how changes in state correlation can affect the forecasting distribution within a Bayesian estimation framework. Section 4 deals with the issues of carrying out MCMC estimation of UC models with a reduced rank covariance matrix. Section 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 6. Section 7 concludes and presents directions for future research.

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 modeling 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 is explored.[4]

2.1 I(1)-UC Models

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

$$y_t = \tau_t + c_t, \tag{1a}$$

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

$$\phi(L)c_t = \varepsilon_t, \tag{1c}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta \\ \rho_{\varepsilon\eta}\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 \end{bmatrix} \right). \tag{1d}$$

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 to as *trend inflation*, which accords well with the Beveridge-Nelson characterization of what represents 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.

When $p = 0$ (i.e., $\phi(L) = 1$), the framework above describes a simple random walk plus noise, or

[3] Data description is deferred to Section 6.1.

[4] Admittedly, the order of integration of inflation is a debatable issue. For example, depending on the sample period, the Dickey-Fuller type of tests suggests one or no unit-root for all inflation measures. Given that the focus here is on out-of-sample performance, we take an agnostic view on pre-testing procedures and produce forecasts based on models that 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.

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 $p = 2$ (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 level fluctuations 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 3 as it constitutes an important motivation to our empirical exercise.

The three structures entertained in this paper 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 applies only 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 et al. (2003) (i.e., MNZ stands for Morley, Nelson and Zivot), 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 $p = 0$): $\tau_t = \tau_{t-1} + \kappa_\tau\varepsilon_t$, $c_t = \varepsilon_t$;
- **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 $p = 2$): $\tau_t = \tau_{t-1} + \kappa_\tau\varepsilon_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$;

[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 $\rho_{\varepsilon\eta}$.

- **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 Morley et al. (2003) can accommodate two variants of MSOE schemes, namely, when $\rho_{\varepsilon\eta} = 0$ and $\rho_{\varepsilon\eta}$ being 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 that 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 $\mathbb{C}\text{ov}(\varepsilon_t, \eta_t^*) = 0$. 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 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:

$$y_t = \tau_t + c_t, \tag{2a}$$

$$\tau_t = \mu_t + \tau_{t-1} + \eta_t, \tag{2b}$$

$$\mu_t = \mu_{t-1} + \zeta_t, \tag{2c}$$

$$\phi(L)c_t = \varepsilon_t, \tag{2d}$$

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

Since τ_t is now specified as a random walk with drift process, such models are useful to model variables that 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

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

$(\Delta = (1-L))$ of y_t and defining $\widetilde{c}_t = (1-L)c_t + \eta_t$, one can re-express the system above as:

$$\Delta y_t = \mu_t + \widetilde{c}_t, \tag{3a}$$

$$\mu_t = \mu_{t-1} + \zeta_t, \tag{3b}$$

$$\phi(L)\widetilde{c}_t = \phi(L)\eta_t + (1-L)\varepsilon_t, \tag{3c}$$

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 $q > 0$; and (b) instead of two innovations, I(2)-UC models can accommodate up to three innovations. Consequently, different correlation structures can be explored. Specifically, we look at three approaches to identifying 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 that bridges the previous two. We construct UC models that 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 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 (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)). Finally, as before, combining the covariance structures above with the different orders of p allows us to construct the following six models:

[7] For inflation, one possible motivation for such identification scheme is to think of changes in trend inflation as mainly reflecting 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.

- **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 a new loading parameter, κ_μ, is introduced in the state equation for μ_t in SSOE variants and can be described in a similar fashion as κ_τ in the I(1)-UC case. Table 1 summarizes all eleven specifications presented in this section.

Table 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.

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 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 $\boldsymbol{\theta}^{UC}$. Letting f denote any probability density function, a density forecast of y_{t+k} is given as follows:[9]

$$\begin{aligned} f(y_{t+k}|\mathbf{y}) &= \int_{\boldsymbol{\Theta}^{UC}} f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{UC})\ f(\theta^{UC}|\mathbf{y}) d\boldsymbol{\theta}^{UC} \\ &= \int_{\boldsymbol{\Theta}^{UC}} f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{UC})\ \frac{f(\mathbf{y}|\theta^{UC}) f(\boldsymbol{\theta}^{UC})}{f(\mathbf{y})} d\boldsymbol{\theta}^{UC}, \end{aligned} \tag{4}$$

where the second term of the integrand in (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 a UC model can alter $\boldsymbol{\Theta}^{UC}$.

To give an example, consider again the local linear trend model shown in Section 2.2. One can readily verify that taking second differences (Δ^2) of the measurement equation in (2a) yields:

$$\Delta^2 y_t = \zeta_t + \eta_t - \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}. \tag{5}$$

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 (0,2,2) process:

$$\Delta^2 y_t = u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2}, \quad u_t \sim \mathcal{N}(0, \sigma^2), \tag{6}$$

[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 ideas underpinning forecasting implications from combining parameter uncertainty and parameter space restrictions can be carried over to discrete and mixed discrete-continuous distributions at the expense of making notation more cumbersome.

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 (5) represent alternative parameterizations of a reduced form ARIMA (0,2,2) model.[10] As a result, the predictive density in (4) can, in principle, also be expressed as:

$$f(y_{t+k}|\mathbf{y}) = \int_{\boldsymbol{\Theta}^{ARIMA}} f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{ARIMA}) \frac{f(\mathbf{y}|\boldsymbol{\theta}^{ARIMA}) f(\boldsymbol{\theta}^{ARIMA})|J|}{f(\mathbf{y})} d\boldsymbol{\theta}^{ARIMA}, \tag{7}$$

where $\boldsymbol{\theta}^{ARIMA} = \{\varphi_1,\ \varphi_2,\ \sigma^2\}$, 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 (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 an invertible and stationary 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 (7) that $f(y_{t+k}|\mathbf{y})$ will not be invariant to such changes.

To illustrate, Figure 1 shows how the SSOE, RSOE and MSOE correlation schemes for the local linear trend generate 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 that 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 a 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 that $f(y_{t+k}|\mathbf{y})$ can account for.[11] Whether or not such constraints are desirable for forecasting is, ultimately, an empirical question.

[10] The canonical representation of an ARIMA (0,2,2) model in (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)).

[11] Derivation of the admissible regions shown in Figure 1 is deferred to Appendix A.3.

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 and stationary ARIMA representation:

Proposition 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 and stationary ARIMA(2,2,3) representation.*

Proof – See Appendix A.1.

4 MCMC Inference of UC Models with Reduced Rank Covariance Matrix

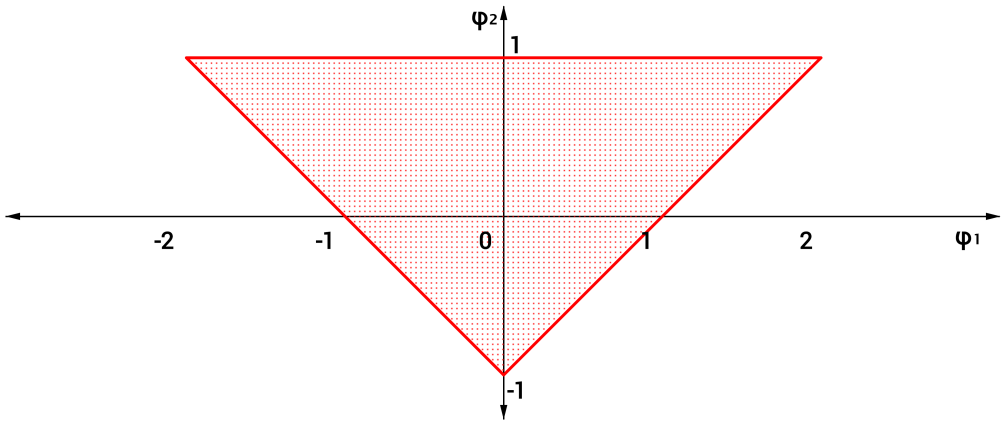
A common approach to estimate UC models (or state space models in general) is to employ MCMC simulation techniques.[12] The usefulness of MCMC sampling in the context of UC models stems from the modular nature of such type of algorithm. In particular, MCMC-based estimation allows one to transform the 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 that 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 paper has not appeared in the literature to date. Therefore, before developing an MCMC algorithm for the models in Section 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. Formal Bayesian estimation of such a model would entail sampling τ_t 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, letting τ_0 and μ_0 denote the initial conditions for τ_t and μ_t

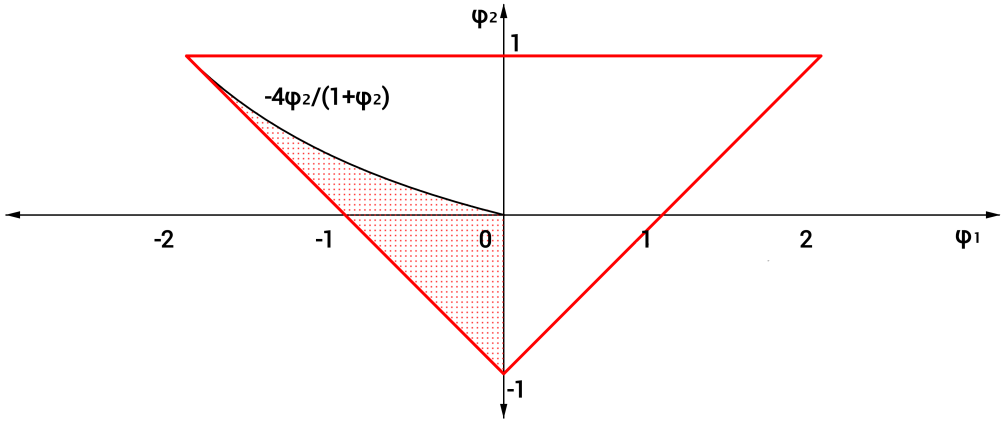
[12] The interested reader is referred to e.g., Koop (2003), Gamerman and Lopes (2006) and the references therein for a detailed textbook treatment on MCMC estimation.

[13] Forbes et al. (2000) and Snyder et al. (2001) address Bayesian estimation of state space models with an 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 that accord well with MCMC sampling, such as stochastic volatility.

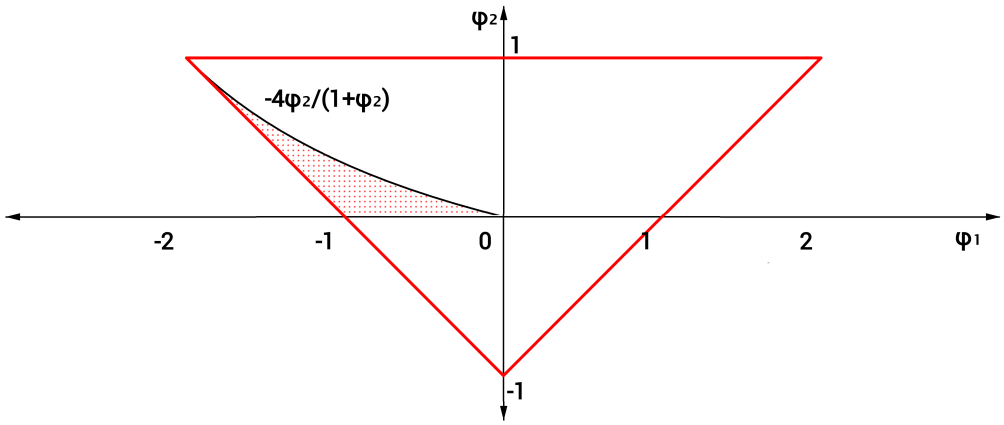
Figure 1: Parameter space restrictions over the invertibility region of reduced-form MA parameters following different correlation structures for the local linear trend model. Dotted area denotes the admissible (or non-constrained) region for each correlation structure.



Local Linear Trend-SSOE



Local Linear Trend-RSOE



Local Linear Trend-MSOE

respectively and given the vector of observations, $\mathbf{y} = (y_1, ..., y_T)'$, a common MCMC algorithm for the local linear trend model can be described as a two-step sampling scheme that involves sequentially drawing from the following conditional posteriors:[14]

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 posterior 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 4.1 below summarizes such considerations:

Proposition 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, 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 A.2.

While Proposition 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 4.1 result from density degeneracies (i.e., probability distributions with zero variance) that 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 that stem from perfectly correlated states can be handled within the Kalman filter by setting the variance-updating step to zero in the forward filtering recursions, which makes the backward smoothing step redundant.[15] 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 that does this by parameterizing UC models in terms of their innovations rather than states.

[14] Initialization of state variables is discussed in Section 5.

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

5 Posterior Analysis

In this section we present an efficient posterior simulator for the UC models discussed in Section 2. In particular, we develop a disturbance-smoothing algorithm that 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 that 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 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 that require simulation of posterior distributions literally billions of times.

For concreteness, consider now the CLARK-MSOE model discussed in Section 2.2. Stacking y_t, τ_t, μ_t and c_t over t for $t = 1, 2..., T$ yields the following matrix representation:

$$\mathbf{y} = \boldsymbol{\tau} + \mathbf{c}, \tag{8}$$

$$\mathbf{H}\boldsymbol{\tau} = \boldsymbol{\iota_0}\tau_0 + \boldsymbol{\mu} + \boldsymbol{\eta} \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\eta), \tag{9}$$

$$\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} \quad \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\zeta), \tag{10}$$

$$\mathbf{H}_\phi\mathbf{c} = \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\varepsilon), \tag{11}$$

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.[16]

Next, we derive a disturbance-based parameterization of the model described in (8)–(11). To do so, notice that by pre-multiplying both sides of (8) by $\mathbf{HH}$ gives:

$$\mathbf{HHy} = \mathbf{HH}\boldsymbol{\tau} + \mathbf{HHc}, \tag{12a}$$

$$\mathbf{HHy} = \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \mathbf{H}\boldsymbol{\mu} + \mathbf{H}\boldsymbol{\eta} + \mathbf{HHc}, \tag{12b}$$

$$(\mathbf{HH})^{-1}\mathbf{H}_\phi\mathbf{HHy} = (\mathbf{HH})^{-1}\mathbf{H}_\phi\left(\mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} + \mathbf{H}\boldsymbol{\eta} + \mathbf{HHc}\right), \tag{12c}$$

$$\widetilde{\mathbf{y}} = \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \tag{12d}$$

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 (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 (12c) makes it easy to verify that (12d) ensues.[17]

Now recall that once we present the measurement equation as in (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 (8)–(11) can be recast as:

$$\widetilde{\mathbf{y}} = \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_\varepsilon), \tag{13}$$

$$\widetilde{\boldsymbol{\eta}} \sim \mathcal{N}(\mathbf{0},\ \mathbf{D}_{\widetilde{\eta}}), \tag{14}$$

$$\widetilde{\boldsymbol{\zeta}} \sim \mathcal{N}(\mathbf{0},\ \mathbf{D}_{\widetilde{\zeta}}), \tag{15}$$

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

[16] 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. For additional approaches to initialize state space models that share similarities with the framework presented here, the reader is referred to De Jong and Chu-Chun-Lin (1994) and Casals and Sotoca (2001).

[17] 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 $\mathbf{H}$ is a lower triangular Toeplitz matrix, so are $\mathbf{HH}$ and $(\mathbf{HH})^{-1}$.

The usefulness of parameterizing UC models in terms of innovations becomes more evident in the case of UC models that contain a covariance matrix with reduced rank. In particular, since rank reduction stems from reducing the number of innovations, the representation in (13)–(15) provides an intuitive way to think of SSOE and RSOE schemes as nested cases within an 'innovations-richer' framework. For example, the SSOE case can be thought of as an MSOE specification where $\mathbf{X_1} = \mathbf{X_2} = \mathbf{0}_{(T\times T)}$ since SSOE models do not require sampling $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\zeta}}$. In fact, all UC variants entertained in this paper can be accommodated into the framework given by (13)–(15) 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.[18] 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 in Table 1 to Appendix A.5.

Obtaining posterior draws for the representation in (13)–(15) can, thus, be summarized as a three-step algorithm that requires sequentially sampling from:

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 $\mathbf{z_{-j}}$ 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: $\boldsymbol{\theta} = \{\sigma_\varepsilon^2,\ \sigma_\eta^2,\ \sigma_\zeta^2,\ \phi_1,\ \phi_2\}$.[19] 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 5.2. We turn next to the discussion of disturbance smoothing.

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

[18] More specifically, adopting general notation as in (13)–(15) enables one to describe posterior moments for all models in terms of common matrix structures.

[19] For notational convenience we exclude $\mathbf{z_0}$ as a conditioning factor in conditional posteriors of the MCMC algorithm since, as previously discussed, initial conditions are predetermined using Winter's approach.

as:

$$f(\mathbf{y}|\mathbf{z},\boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_\varepsilon|^{-\frac{1}{2}} \exp\left(-\frac{(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}})'\boldsymbol{\Sigma}_\varepsilon^{-1}(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}})}{2}\right), \tag{16}$$

and

$$f(\widetilde{\boldsymbol{\eta}}|\boldsymbol{\theta}) \propto |\mathbf{D}_{\tilde{\eta}}|^{-\frac{1}{2}} \exp\left(-\frac{\widetilde{\boldsymbol{\eta}}'\mathbf{D}_{\tilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}}}{2}\right), \tag{17}$$

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

Next, using Bayes' rule to combine (16) with (17) and applying standard regression results (e.g., Koop et al. (2007)) yields:

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

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 $\overline{\mathbf{D}}_{\tilde{\eta}} = \left(\mathbf{X}_\mathbf{2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)^{-1}$ and $\bar{\mathbf{d}}_{\tilde{\eta}} = \overline{\mathbf{D}}_{\tilde{\eta}}\mathbf{X}_\mathbf{2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\widetilde{\mathbf{y}}_{\tilde{\eta}}$. Now, remember that we defined $\mathbf{X_2} = \mathbf{H}_\phi$ and $\mathbf{D}_{\tilde{\eta}} = \mathbf{H}^{-1}\boldsymbol{\Sigma}_\eta\mathbf{H}'^{-1}$. 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 pentadiagonal 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 $\overline{\mathbf{D}}_{\tilde{\eta}} = \left(\mathbf{X}_\mathbf{2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)^{-1}$ 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 $\left(\mathbf{X}_\mathbf{2}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)^{-1}$ via brute-force inversion (see e.g., Golub and Van Loan (1983) p. 156).

To illustrate how we adapt the algorithm in Chan and Jeliazkov (2009) 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., $\mathbf{C} \setminus \mathbf{b} = \mathbf{C}^{-1}\mathbf{b}$. 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:

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

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) that 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:

$$\overline{\mathbf{d}}_{\widetilde{\eta}} = \mathbf{C}'^{-1}\left(\mathbf{C}^{-1}\left(\mathbf{X}_{\mathbf{2}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\widetilde{\mathbf{y}}_{\widetilde{\eta}}\right)\right) = (\mathbf{C}\mathbf{C}')^{-1}\left(\mathbf{X}_{\mathbf{2}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\widetilde{\mathbf{y}}_{\widetilde{\eta}}\right) = \overline{\mathbf{D}}_{\widetilde{\eta}}^{-1}\left(\mathbf{X}_{\mathbf{2}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\widetilde{\mathbf{y}}_{\widetilde{\eta}}\right).$$

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 (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 $\overline{\mathbf{D}}_{\widetilde{\zeta}} = \left(\mathbf{X}_{\mathbf{1}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X}_{\mathbf{1}} + \mathbf{D}_{\widetilde{\zeta}}^{-1}\right)^{-1}$, $\overline{\mathbf{d}}_{\widetilde{\zeta}} = \overline{\mathbf{D}}_{\widetilde{\zeta}}\mathbf{X}_{\mathbf{1}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\widetilde{\mathbf{y}}_{\widetilde{\zeta}}$ and $\widetilde{\mathbf{y}}_{\widetilde{\zeta}} = \widetilde{\mathbf{y}} - \mathbf{X}_{\mathbf{0}}\mathbf{z}_{\mathbf{0}} - \mathbf{X}_{\mathbf{2}}\widetilde{\boldsymbol{\eta}}$. Note that $\mathbf{X}_{\mathbf{1}} = \mathbf{H}_{\phi}$, $\mathbf{D}_{\widetilde{\zeta}}^{-1} = (\mathbf{H}\mathbf{H})'\boldsymbol{\Sigma}_{\zeta}^{-1}(\mathbf{H}\mathbf{H})$ and $\mathbf{X}_{\mathbf{1}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X}_{\mathbf{1}}$ are all sparse matrices. Therefore, draws of $\widetilde{\boldsymbol{\zeta}} = (\mathbf{H}\mathbf{H})^{-1}\boldsymbol{\zeta}$ can also be quickly obtained adjusting the four $\mathcal{O}(T)$ steps described above to the conditional posterior mean and variance of $\widetilde{\boldsymbol{\zeta}}|\mathbf{y}, \mathbf{z}_{\_\widetilde{\zeta}}, \boldsymbol{\theta}$ (i.e., $\overline{\mathbf{d}}_{\widetilde{\zeta}}$ and $\overline{\mathbf{D}}_{\widetilde{\zeta}}$ respectively). Of course, given $\widetilde{\boldsymbol{\zeta}}$, setting $\boldsymbol{\zeta} = \mathbf{H}\mathbf{H}\widetilde{\boldsymbol{\zeta}}$ allows one to recover $\boldsymbol{\zeta}$.

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 appear only in UC models containing two or more correlated states. Similarly, ϕ_1 and ϕ_2 need to be sampled only 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. Moreover, let $\boldsymbol{\theta_{-\sigma^2}}$ denote any specification-consistent subset of $\boldsymbol{\theta}$ 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:[20]

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

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

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 that 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. In addition to Proposition 3.1, the exact invertibility ($A_{\boldsymbol{\psi}}$) and stationary ($A_{\boldsymbol{\phi}}$) conditions for the UC models adopted in this paper are discussed in Appendix 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:

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

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

[20] Prior hyperparameters are discussed in Section 6.1

Next, draws of the loading and autoregressive parameters require simulating $f(\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma}^2)$. The latter, however, is not of a known form that 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 (16) with a joint prior density $f(\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2})$ yields:

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

where $\widetilde{\mathbf{y}}_* = \widetilde{\mathbf{y}} - \mathbf{X_0}\mathbf{z_0} - \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}}$. The expression above can be quickly evaluated using sparse routines implemented in most statistical packages.[21] More specifically, draws from $f(\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma}^2)$ are obtained using an independence-chain Metropolis-Hastings step (see, e.g., Tierney (1994)) with proposal density given by $\mathcal{N}\left(\hat{\boldsymbol{\theta}}_{-\boldsymbol{\sigma}^2},\ \mathbf{G}^{-1}\right)$ where a Newton-Raphson method is adopted to numerically compute the mode ($\hat{\boldsymbol{\theta}}_{-\boldsymbol{\sigma}^2}$) and the negative Hessian ($\mathbf{G}^{-1}$) evaluated at the mode of $f(\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma}^2)$.

6 Evaluation

In this section we empirically evaluate the effects on forecasting performance that stem from allowing for different state correlation structures. Even though our focus is on forecasting, we also present (i) results for trend inflation measures that arise from different correlation structures and (ii) computational efficiency results for the MCMC algorithm developed in Section 5.

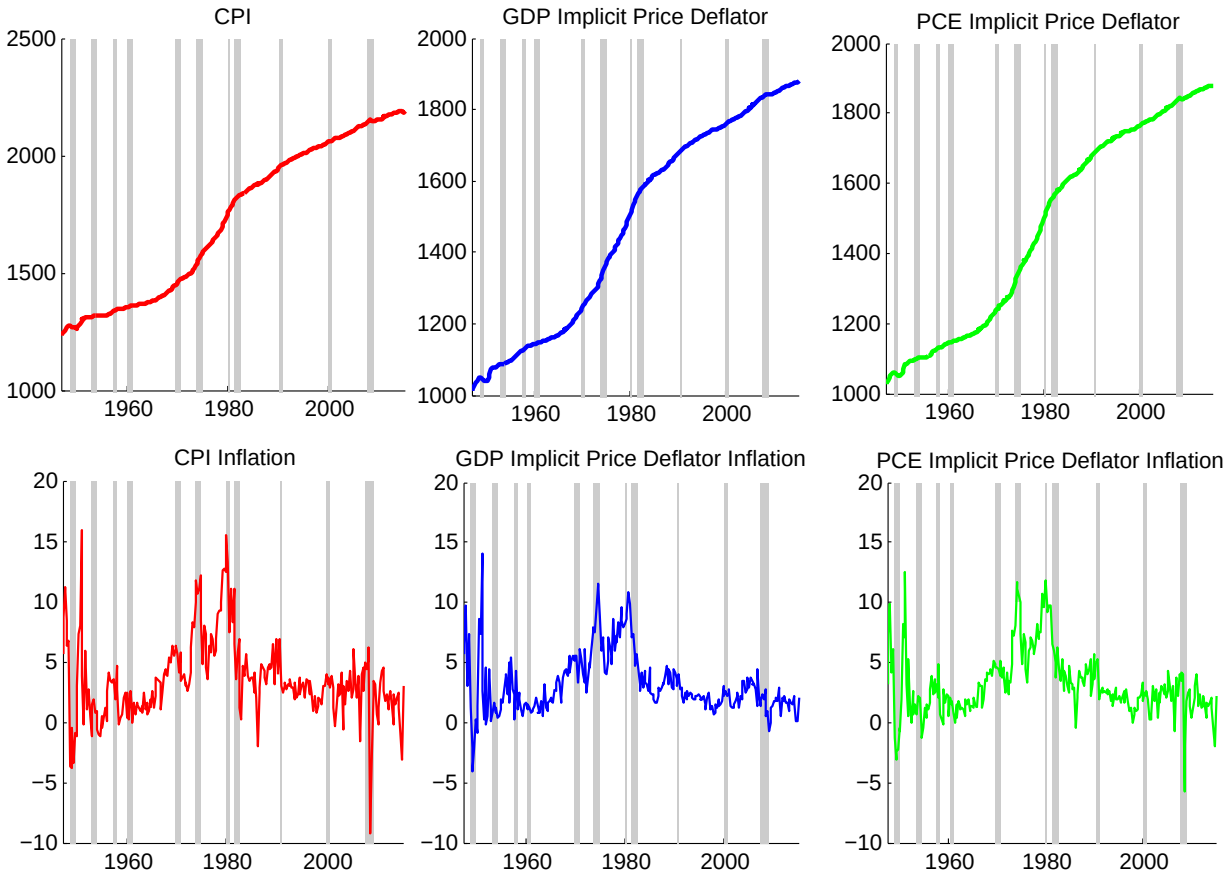
6.1 Data and Priors

Our data set consists of quarterly annualized series for CPI and the implicit price deflators of real GDP and personal consumption expenditure (PCE). All series range from 1947Q1 to 2015Q2. Remember that for I(1)-UC and I(2)-UC models, observations correspond to the inflation rate and (log) price level respectively. Inflation series are constructed using the first difference in logs of price levels. Figure 2 shows the data used in our empirical exercise.

Priors are selected to balance three criteria: (i) to facilitate comparison across models; (ii) to be relatively uninformative; and (iii) to follow recommendations from previous studies. For example, in keeping with previous forecasting studies (e.g., Stock and Watson (2007), Chan (2013) and Clark

[21] Depending on the UC model specification, parameters in $\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2}$ show up in different parts of the measurement equation in (13). In the CLARK-MSOE model, for example, recall from our previous discussion that $\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2} = \{\phi_1,\ \phi_2\}$, $\mathbf{X_0} = ((\mathbf{HH})^{-1}\mathbf{H}_\phi\mathbf{H}\boldsymbol{\iota}_0\ \ (\mathbf{HH})^{-1}\mathbf{H}_\phi\boldsymbol{\iota}_0)$, and $\mathbf{X_1} = \mathbf{X_2} = \mathbf{H}_\phi$. Nonetheless, regardless of the specification, all such matrices are banded. Once again, the interested reader is referred to technical Appendix A.5, for a detailed description of the exact structures underlying the matrices in the disturbance-based parameterization of all UC models employed in this paper.

Figure 2: U.S. Quarterly Measures of Annualized Price Level and Inflation from 1947Q1 to 2015Q2



Note: Shaded regions indicate recessions as recorded by the NBER

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 a UC representation that assigns quite small conditional variance of the trend level state (i.e., τ_t).[22] 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 summarizes prior densities for each model.

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

Table 2: Priors

Model	σ_ε^2	σ_η^2	σ_ζ^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	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	–	–
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	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\psi\in A_\phi)}$
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	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9^{-10^6})$	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$

6.2 The Forecasting Algorithm

We now use all models listed in Table 1 to carry out a recursive forecasting exercise for the series in Figure 2.[23] Inflation 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 predictive log-scores.

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 that, under the null hypothesis, postulates equivalent forecasting accuracy between competing models. Such a test is applied for both RMSFE and predictive log-score results.[24] To control for serial correlation in forecast errors, as in Clark and Doh (2014), standard errors of t-statistics for the Diebold and Mariano (1995) test are computed using a heteroskedasticity and autocorrelation-consistent with pre-whitened quadratic spectral estimator.

To generate inflation forecasts we conduct a predictive simulation exercise along the lines of Cogley et al. (2005) adapted to the disturbance-based parametrization discussed in Section 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, using standard results for conditional probability, a k-step-ahead predictive density for Δy_{t+k} can be expressed as:

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

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

[24] 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.

where

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

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 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, then for every *r-th* MCMC draw of $\mathcal{F}$ (i.e., $\mathcal{F}^{(r)}$) 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:

Given information up to time t we (sequentially):
Step 1: Simulate parameters ($\boldsymbol{\theta}$) and the vector of innovations ($\mathbf{z_{1:t}}$) from the posterior density, $f(\mathbf{z_{1:t}}, \boldsymbol{\theta}|\mathbf{y_{1:t}})$, at *each r-th* iteration of the MCMC algorithm discussed in Section 5.
Step 2: Simulate innovations in $\mathbf{z_{t+1:t+k}}$ from the normal predictive density, $f(\mathbf{z_{t+1:t+k}}|\mathbf{y_{1:t}}, \mathbf{z_{1:t}}, \boldsymbol{\theta})$, using second moment parameters obtained in Step 1.
Step 3: Simulate forecasts, $\boldsymbol{\Delta}\mathbf{y_{t+1:t+k}} = \{\Delta y_{t+1}, ..., \Delta y_{t+k}\}$, from the normal predictive likelihood, $f(\boldsymbol{\Delta}\mathbf{y_{t+1:t+k}}|\mathbf{y_{1:t}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta}) = \prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta})$, using innovations and parameters generated in Steps 1 and 2.

Once the algorithm above is performed R times, post burn-in draws of Δy_{t+k} (stored in Step 3) provide an approximation of the predictive density $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$. Such draws can then be used to construct point and density forecasts that account for data up to point t. Moving forward one period and repeating Steps 1 through 3 R times produces a new set of (point and density) forecasts, except now using information up to point $t+1$. We keep iterating in this fashion until point $T-1$ to generate RMSFE and predictive log-scores results reported below.

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. Remember from Section 2.2 that $\mu_t = \mu_{t-1} + \zeta_t$, $\widetilde{c}_t = (1-L)c_t + \eta_t$ and $\phi(L)c_t = \varepsilon_t$, then pre-multiplying both sides of the measurement equation in (3a) by $\phi(L)$ and moving it $t+k$ periods forward gives:

$$\begin{aligned}\phi(L)\Delta y_{t+k} &= \phi(L)\mu_{t+k} + \phi(L)\widetilde{c}_{t+k} \\ &= \phi(1)\mu_0 + \phi(L)\sum_{j=1}^{t+k}\zeta_j + \phi(L)\eta_{t+k} + (1-L)\varepsilon_{t+k}.\end{aligned}$$

The second expression above indicates that (point and density) forecasts for Δy_{t+k} can be obtained from an ARIMA process written in terms of UC model innovations. In particular, $\{\eta_j,\ \zeta_j\}_{j=1}^{t}$ and $\phi(L)$ are obtained from our MCMC sampler (Step 1);[25] $\{\eta_{t+s},\ \zeta_{t+s}\}_{s=1}^{k}$ are simulated from $\eta_{t+s} \sim \mathcal{N}(0,\ \sigma_\eta^2)$ and $\zeta_{t+s} \sim \mathcal{N}(0, \sigma_\zeta^2)$ (Step 2). Therefore, Δy_{t+s} for $s = 1, ..., k$ can be generated sequentially by simulating $\varepsilon_{t+s} \sim \mathcal{N}(0,\ \sigma_\varepsilon^2)$ and treating $\{\eta_j,\ \zeta_j\}_{j=1}^{t+s}$ as predetermined variables (Step 3).

It is easy to see that analogous algebraic steps can be applied to all other UC models to obtain a predictive parameterization in terms of innovations. Of course, in-sample correlation structure between states is preserved out-of-sample. In other words, only innovations entering the model are simulated in Step 2 of the predictive sampler described above.

6.3 Results: Point Forecasts

To compare point forecast accuracy across models we compute the RMSFE for each k-step-ahead prediction defined as

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

where Δy_{t+k}^o denotes the actual value of inflation that is known at time $t+k$, while t_0 and $T-k$ denote the first and last forecast generated respectively.[26] In particular, we set t_0 =1971Q1 and T =2015Q2. An estimate of the mean of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$, $\widehat{\mathbb{E}}(\Delta y_{t+k}|\mathbf{y_{1:t}})$, is constructed by averaging over R draws of Δy_{t+k} generated using the predictive simulation algorithm described in Section

[25] Again, initial conditions $(\tau_0,\ \mu_0)$, as discussed previously, are obtained using the exponential smoothing approach of Winters (1960).

[26] To keep the discussion in Sections 6.3 and 6.4 consistent with the presentation of the forecasting algorithm in Section 6.2 we refer to Δy_{t+k} as an *inflation* forecast. Of course, for forecasts based on I(1)-UC models there is no need to take first differences of y_t to construct the RMSFE and predictive log-score metrics since such specifications model y_t as inflation directly.

6.2.[27]

To make comparison of forecast performance across models easier, Table 3 reports RMSFE results relative to the local level-MSOE model. Specifically, entries in Table 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 that outperform the local level-MSOE model. Also, numbers in bold denote the best performing model for a specific forecast horizon.

Overall, results in Table 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-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 for long-run (point) forecasting. Such a result is likely to reflect the usefulness of smoother measures of trend inflation (see Figure 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 that are statistically significant at the 5% level relative to the orthogonal local level model. Also, I(2)-UC models that 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 and their orthogonal counterparts at two-, three- and four-year-ahead forecast horizons.

To summarize, results in Table 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.

[27] 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.

Table 3: Relative RMSFEs for U.S. 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.

6.4 Results: Density Forecasts

As seen in Table 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 that 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 that incorporate uncertainty around the prediction location, then the predictive accuracy between such densities would differ commensurate with their difference in variances.

Now, recall from our discussion in Section 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 a UC model accommodates, as measured in terms of the restrictions a UC model specification imposes over its corresponding reduced form ARIMA parameter space. Since changes in the correlation structure alter the parameter space over that $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$ is defined, then forecast metrics that incorporate information within the full support of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$ are, in principle, more likely to reflect implications between out-of-sample performance and the error correlation structure that may be silent in a point forecasting framework.

A natural candidate to evaluate density forecasts is the sum of log predictive likelihoods:[28]

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

where $\widehat{f}(\Delta y^o_{t+k}|\mathbf{y_{1:t}})$ is computed using the predictive algorithm in Section 6.2. The rationale underlying the metric above is that if actual inflation outcome, $\Delta y^o_{t+k}|\mathbf{y_{1:t}}$, is unlikely under the density forecast, $f(\Delta y^o_{t+k}|\mathbf{y_{1:t}})$, then the value of $\log(f(\Delta y^o_{t+k}|\mathbf{y_{1:t}}))$ will be small. 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 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 4 reinforce 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

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

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.

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.

Figure 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 a result most likely reflects the fact that latent components in the SSOE case are recovered from a unique source of randomness that 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 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 that are quite smooth. As such, to the extent that smoother measures of trend inflation are preferable for policy analysis (see, e.g., Castelnuovo et al. (2014)), RSOE UC models can be perceived to provide more suitable correlation structures. From a statistical viewpoint (or in terms of the signal to noise ratio), such smoothness suggests that treating unexplained 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) reflects only strong signals from changes in the price level.

6.6 Computation Efficiency

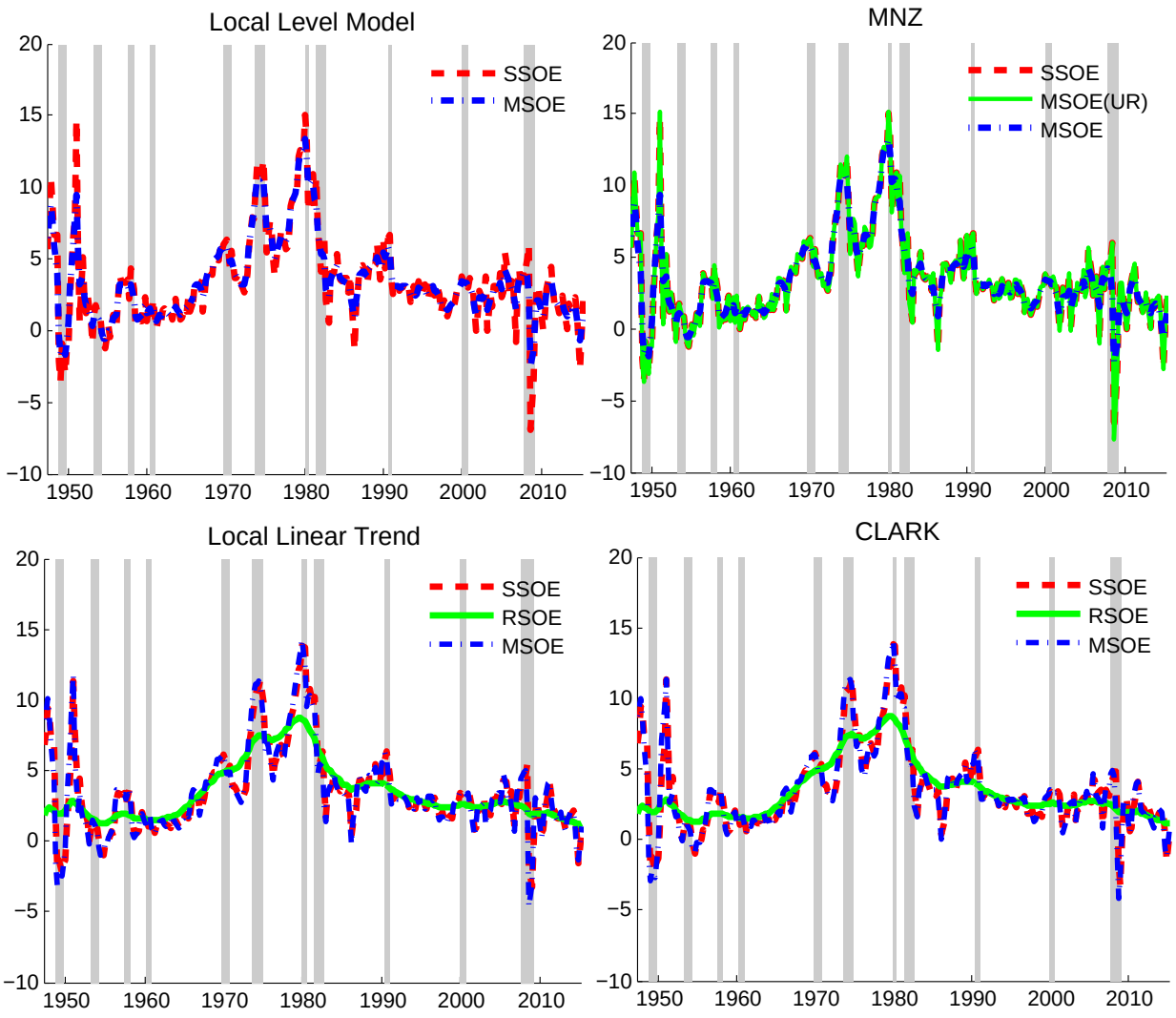
To assess the performance of an MCMC sampler, a common approach is to verify its mixing properties. In this sense, an MCMC algorithm with good mixing properties is one that allows the researcher to interpret parameter draws as independent realizations of a random variable. Consequently, an algorithm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

Table 4: Sum of Log Predictive Likelihoods for U.S. 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

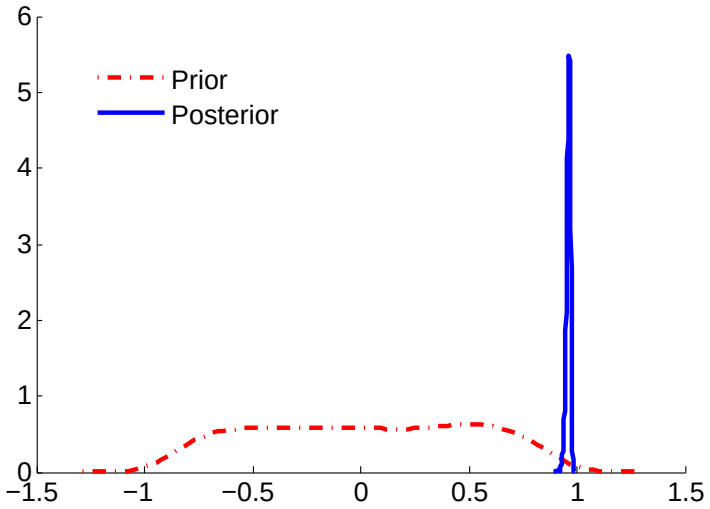
* indicates superior forecast performance relative to the local level-MSOE model at the 5% level of significance using a Diebold and Mariano (1995) test for equivalence in squared forecast errors.

Figure 3: U.S. Quarterly Measures of Annualized Trend (CPI) Inflation



Note: Shaded regions indicate recessions as recorded by the NBER

Figure 4: Kernel Smoothed Density for the Correlation between τ_t and c_t under MNZ-MSOE(UR) Model



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 particularly be an issue for I(2)-UC models since the conditional variance of y_t is decomposed into three stochastic processes with (at least) one of them typically exhibiting near-zero 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.[29] Therefore, although UC models are reparameterized in this paper to address rank-reduction issues, a natural question is whether or not the disturbance-based parameterization combined with precision sampling techniques – discussed in Section 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 should be one. As a rule of thumb, inefficiency factors around twenty based on the metric above are typically interpreted as an indication of fast mixing.[30] Table 5 reports the inefficiency factors for parameter draws associated with each of the eleven models. Figure 5 and Figure 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 summarize inefficiency factor results. In particular, the middle line denotes the median inefficiency factor based on a sample of inefficiency factors for each element in $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ constructed using post burn-in MCMC draws. Similarly, lower and upper lines respectively represent the 25 and 75 percentiles, while whiskers extend to the maximum and minimum inefficiency factors. All in all, our results below suggest that the algorithm developed in Section 5 is quite efficient in terms of generating parameters and innovations draws that are not strongly autocorrelated.

[29] 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.

[30] 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 5: Inefficiency Factors for Disturbance Smoothing (I(1)-UC Models)

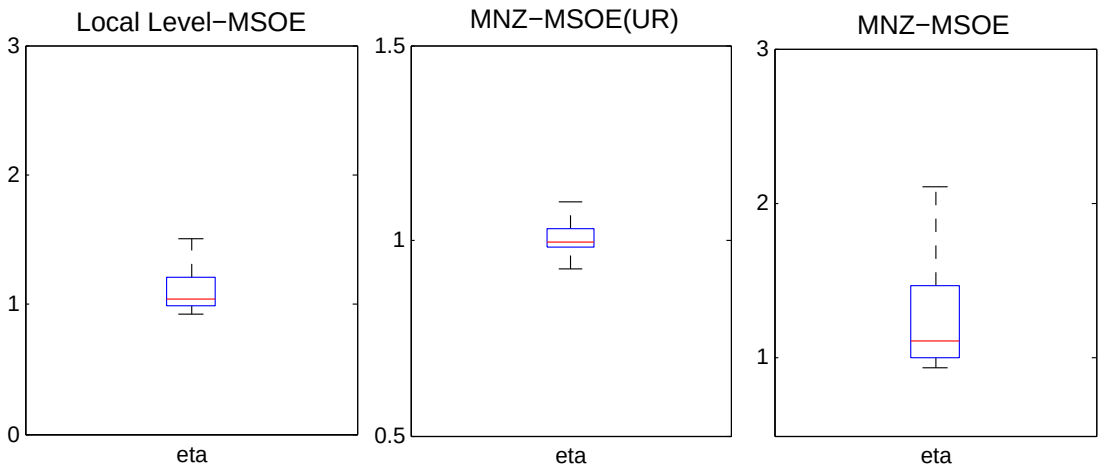


Figure 6: Inefficiency Factors for Disturbance Smoothing (I(2)-UC Models)

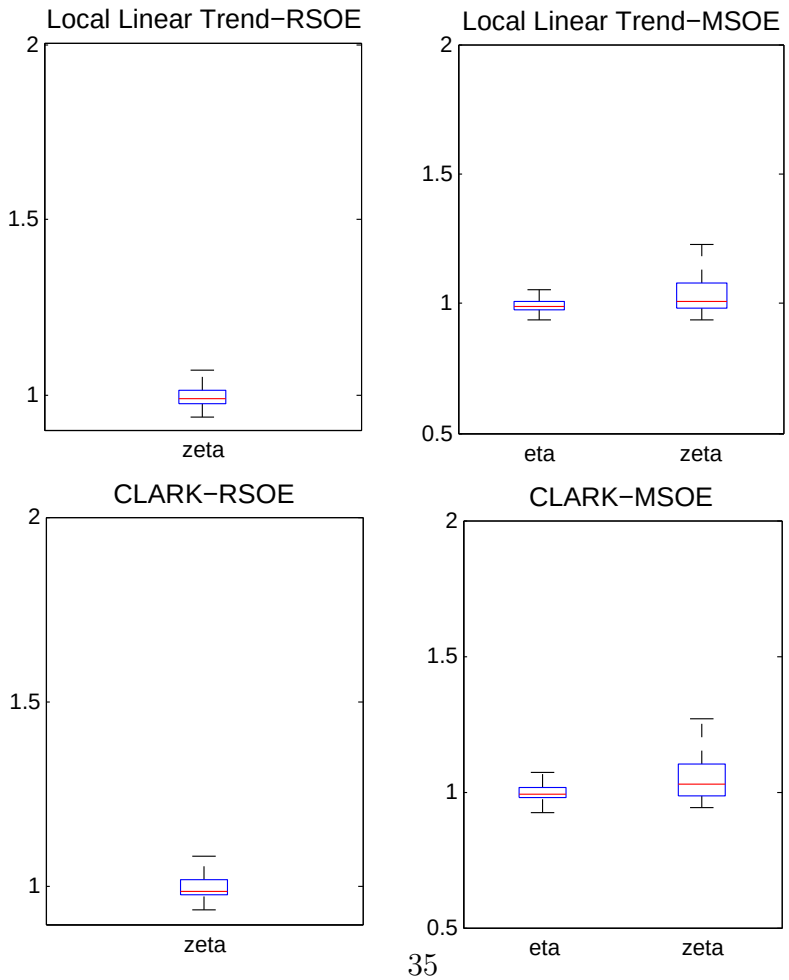


Table 5: Inefficiency Factors for Parameter Sampling

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

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 $\boldsymbol{\theta}_{\boldsymbol{-\sigma^2}}|\widetilde{\mathbf{y}}, \mathbf{z}, \boldsymbol{\sigma^2}$. In terms of computational speed, estimation of all eleven UC models — where 10000 draws are sampled for each model — takes less than 180 seconds.

7 Concluding Remarks and Extensions

In this paper 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 models. 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 with regard 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 that 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 a variant also generates smooth measures of trend inflation, which is typically a desirable feature for policy analysis.

Another contribution from this paper 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 to formulate multivariate versions of the UC models entertained in this study.

A Appendix

A.1 Proof of Proposition 3.1

To prove part 3.1(a), recall first that the local linear trend-MSOE model in section 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:

$$\Delta^2 y_t = \zeta_t + \left[\kappa_\tau(1-L) + 1 - 2L + L^2\right]\varepsilon_t. \tag{18}$$

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 (18) denotes the sum of two independent MA processes, namely an MA(0) (i.e., ζ_t) and an 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 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 an MA(3) reduced form representation. Note that the AR polynomial in the left-hand side of the representation above is identical to an AR polynomial of an otherwise reduced form ARIMA representation. Therefore, ensuring that the roots of $\phi(L) = (1 - \phi_1 L - \phi_2 L^2)$ lie outside the unit circle ensures that the CLARK-RSOE model has a stationary ARIMA (2,2,3) representation.

To ensure the resulting MA polynomial is invertible, we 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 $[\kappa_\tau \phi(L)(1-L) + (1-L)^2]\,\varepsilon_t$, do not share such type of roots.

A.2 Proof of Proposition 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 4.1. To do so, we prove Proposition 4.1 by contradiction. Since all models entertained in this paper 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 that 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 **A** being nonsingular. Since **A** cannot be inverted, it is rank-deficient.

A.3 Derivation of the Restrictions Over the MA Parameter Space Shown in Figure 1

In this section we show how to obtain the nonlinear restrictions over the MA parameter space as shown in Figure 1. To that end, for the MSOE and RSOE variants we take a similar approach to that 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 (6) and its corresponding representation given in (5), i.e., in terms of the local linear trend model parameters. Thus, we have:

$$\rho(s) = \begin{cases} \frac{\varphi_1(1+\varphi_2)}{1+\varphi_1^2+\varphi_2^2} = -\frac{\sigma_\eta^2+4\sigma_\varepsilon^2+4\sigma_{\eta\varepsilon}+2\sigma_{\zeta\varepsilon}+\sigma_{\eta\zeta}}{2\sigma_\eta^2+6\sigma_\varepsilon^2+\sigma_\zeta^2+2\sigma_{\eta\zeta}+6\sigma_{\eta\varepsilon}+2\sigma_{\zeta\varepsilon}} & \text{if } s=1, \\ \frac{\varphi_2}{1+\varphi_1^2+\varphi_2^2} = \frac{\sigma_\varepsilon^2+\sigma_{\eta\varepsilon}+\sigma_{\zeta\varepsilon}}{2\sigma_\eta^2+6\sigma_\varepsilon^2+\sigma_\zeta^2+2\sigma_{\eta\zeta}+6\sigma_{\eta\varepsilon}+2\sigma_{\zeta\varepsilon}} & \text{if } s=2, \\ 0 \quad \text{if } s \geq 3, \end{cases} \tag{19}$$

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 (19) and rearranging terms yields:

$$\varphi_1 = -(4+q)\frac{\varphi_2}{1+\varphi_2}, \tag{20a}$$

$$\varphi_2 = g_1\sigma_\varepsilon^2, \tag{20b}$$

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 (20b) it is to 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 (20a) 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 (20a) 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 (20a) for a wide range of values

of $-(4+q) \in (-\infty, -4)$ to construct the space denoted by all such hyperbolas that intersect with the $\varphi_1 < 0$ and $\varphi_2 > 0$ quadrant and 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 1.

• *Local Linear Trend-RSOE*

First, recall that the 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 (19) and rearranging terms yields:

$$\varphi_1 = -\frac{(2+\kappa_\tau)^2}{1+\kappa_\tau} \frac{\varphi_2}{1+\varphi_2}, \tag{21a}$$

$$\varphi_2 = g_2 (1+\kappa_\tau) \sigma_\varepsilon^2, \tag{21b}$$

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 (21b) 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 (21a), 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 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 a model we set: $\eta_t = \kappa_\tau \varepsilon_t$ and $\zeta_t = \kappa_\mu \varepsilon_t$. Plugging these into (5) yields:

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

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 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.

A.4 Invertibility and Stationarity for SSOE and MSOE Models

Stationarity and invertibility conditions for MSOE models are straightforward. In particular, positive definiteness of the covariance matrix for the error terms in the MSOE case ensures that the UC model has a corresponding invertible reduced form ARIMA representation (see e.g., Teräsvirta (1977) and Harvey (2006)). Stationarity conditions follow — as in Appendix A.1 — from noting the equivalence of the AR polynomial for both the reduced form and UC innovations-based ARIMA representations. Therefore, the stationarity conditions presented below for SSOE models are equivalent for MSOE 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 differences of the measurement equation of the local level-SSOE model yields:

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

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:

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

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:

$$\phi(L)\Delta y_t = (1+\kappa_\tau)\varepsilon_t - (\kappa_\tau\phi_1 + 1)\varepsilon_{t-1} - \kappa_\tau\phi_2\varepsilon_{t-2}, \tag{22}$$

$$\phi(L)\Delta y_t = u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2}, \tag{23}$$

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 (22) and (23) 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:

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

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 the ARIMA(2,2,3) process above is the same regardless of whether 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, $\phi(L)$.

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

In this section we show, as claimed in Section 5, that all eleven UC models entertained in this paper 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 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 (12a)–(12d), i.e., we apply the commutative property of lower triangular Toeplitz matrices to construct a disturbance-based parameterization that 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 (12b). For I(1)-UC models, a slightly modified version of (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 and letting $\setminus$ denote the generalized inverse operator, then adjusting each UC model according to its correlation structure, as discussed in Section 2, entails straightforward algebraic manipulations presented below.

I(2)-UC Models

•*CLARK-RSOE*

$$\begin{aligned}
\mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HHc}, \\
\mathbf{H}_\phi\mathbf{HHy} &= \mathbf{H}_\phi\mathbf{H}\boldsymbol{\iota_0}\tau_0 + \mathbf{H}_\phi\boldsymbol{\iota_0}\mu_0 + \mathbf{H}_\phi\boldsymbol{\zeta} + (\kappa_\tau\mathbf{H}_\phi\mathbf{H} + \mathbf{HH})\,\boldsymbol{\varepsilon}, \\
\mathbf{H}_\phi\mathbf{HH}(\mathbf{A} \setminus \mathbf{y}) &= (\mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \;\; \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\iota_0}))\,(\tau_0 \;\; \mu_0)' + \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}, \\
\widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \kappa_\tau \mathbf{H}_\phi \mathbf{H} + \mathbf{HH}, \\ \mathbf{X_0} & = \left(\mathbf{H}_\phi \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \;\; \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\iota_0})\right), \\ \mathbf{X_1} & = \mathbf{H}_\phi, \\ \mathbf{X_2} & = \mathbf{0}_{T\times T}, \\ \mathbf{z_0} & = (\tau_0 \;\; \mu_0)', \\ \widetilde{\mathbf{y}} & = \mathbf{H}_\phi \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}), \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{A} \setminus \boldsymbol{\zeta}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{0}_{T\times 1}. \end{cases}$$

•*CLARK-SSOE*

$$\begin{aligned} \mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \kappa_\mu \boldsymbol{\varepsilon} + \kappa_\tau \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HHc}, \\ \mathbf{H}_\phi \mathbf{HHy} &= \mathbf{H}_\phi \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \mathbf{H}_\phi \boldsymbol{\iota_0}\mu_0 + \left(\kappa_\mu \mathbf{H}_\phi + \kappa_\tau \mathbf{H}_\phi \mathbf{H} + \mathbf{HH}\right)\boldsymbol{\varepsilon}, \\ \mathbf{H}_\phi \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}) &= \left(\mathbf{H}_\phi \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \quad \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\iota_0})\right)(\tau_0 \;\; \mu_0)' + \boldsymbol{\varepsilon}, \\ \widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \kappa_\mu \mathbf{H}_\phi + \kappa_\tau \mathbf{H}_\phi \mathbf{H} + \mathbf{HH}, \\ \mathbf{X_0} & = \left(\mathbf{H}_\phi \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \;\; \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\iota_0})\right), \\ \mathbf{X_1} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_2} & = \mathbf{0}_{T\times T}, \\ \mathbf{z_0} & = (\tau_0 \;\; \mu_0)', \\ \widetilde{\mathbf{y}} & = \mathbf{H}_\phi \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}), \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{0}_{T\times 1}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{0}_{T\times 1}. \end{cases}$$

•*Local Linear Trend-MSOE*

$$\begin{aligned} \mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} + \mathbf{H}\boldsymbol{\eta} + \mathbf{HH}\boldsymbol{\varepsilon}, \\ \mathbf{y} &= \left(\mathbf{H} \setminus \boldsymbol{\iota_0} \quad (\mathbf{HH}) \setminus \boldsymbol{\iota}_0\right)(\tau_0 \;\; \mu_0)' + (\mathbf{HH})\backslash\boldsymbol{\zeta} + \mathbf{H}\backslash\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \\ \widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_0} & = (\mathbf{H} \setminus \boldsymbol{\iota_0} \;\; (\mathbf{HH}) \setminus \boldsymbol{\iota}_0), \\ \mathbf{X_1} & = \mathbf{I}_{(T\times T)}, \\ \mathbf{X_2} & = \mathbf{I}_{(T\times T)}, \\ \mathbf{z_0} & = (\tau_0 \;\; \mu_0)', \\ \widetilde{\mathbf{y}} & = \mathbf{y}, \\ \widetilde{\boldsymbol{\zeta}} & = (\mathbf{HH})\backslash\boldsymbol{\zeta}, \\ \widetilde{\boldsymbol{\eta}} & = (\mathbf{H})\backslash\boldsymbol{\eta}. \end{cases}$$

•*Local Linear Trend-RSOE*

$$\begin{aligned} \mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HH}\boldsymbol{\varepsilon}, \\ \mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \boldsymbol{\zeta} + (\kappa_\tau\mathbf{H} + \mathbf{HH})\,\boldsymbol{\varepsilon}, \\ \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}) &= (\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \;\; (\mathbf{A} \setminus \boldsymbol{\iota_0}))\,(\tau_0 \;\; \mu_0)' + \mathbf{A} \setminus \boldsymbol{\zeta} + \boldsymbol{\varepsilon}, \\ \widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \kappa_\tau\mathbf{H} + \mathbf{HH}, \\ \mathbf{X_0} & = (\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \;\; (\mathbf{A} \setminus \boldsymbol{\iota_0})), \\ \mathbf{X_1} & = \mathbf{I}_{(T\times T)}, \\ \mathbf{X_2} & = \mathbf{0}_{T\times T}, \\ \mathbf{z_0} & = (\tau_0 \;\; \mu_0)', \\ \widetilde{\mathbf{y}} & = \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}), \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{A} \setminus \boldsymbol{\zeta}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{0}_{T\times 1}. \end{cases}$$

•*Local Linear Trend-SSOE*

$$\begin{aligned} \mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \kappa_\mu\boldsymbol{\varepsilon} + \kappa_\tau\mathbf{H}\boldsymbol{\varepsilon} + \mathbf{HHc}, \\ \mathbf{HHy} &= \mathbf{H}\boldsymbol{\iota_0}\tau_0 + \boldsymbol{\iota_0}\mu_0 + \left(\kappa_\mu\mathbf{I}_{(T\times T)} + \kappa_\tau\mathbf{H} + \mathbf{HH}\right)\boldsymbol{\varepsilon}, \\ \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}) &= (\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \quad (\mathbf{A} \setminus \boldsymbol{\iota_0}))\,(\tau_0 \;\; \mu_0)' + \boldsymbol{\varepsilon}, \\ \widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \kappa_\mu \mathbf{I}_{(T\times T)} + \kappa_\tau \mathbf{H} + \mathbf{HH}, \\ \mathbf{X_0} & = \left(\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}) \;\; (\mathbf{A} \setminus \boldsymbol{\iota_0})\right), \\ \mathbf{X_1} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_2} & = \mathbf{0}_{T\times T}, \\ \mathbf{z_0} & = (\tau_0 \;\; \mu_0)', \\ \widetilde{\mathbf{y}} & = \mathbf{HH}(\mathbf{A} \setminus \mathbf{y}), \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{0}_{T\times 1}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{0}_{T\times 1}. \end{cases}$$

I(1)-UC Models

•*MNZ-MSOE*

$$\begin{aligned} \mathbf{Hy} &= \boldsymbol{\iota_0}\tau_0 + \boldsymbol{\eta} + \mathbf{Hc}, \\ \mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi\boldsymbol{\iota_0}\tau_0 + \mathbf{H}_\phi\boldsymbol{\eta} + \mathbf{H}\boldsymbol{\varepsilon}, \\ \mathbf{H^{-1}H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\iota_0})\tau_0 + \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\eta}) + \boldsymbol{\varepsilon}, \\ \widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon}, \end{aligned}$$

$$\text{s.t.} \begin{cases} \mathbf{A} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_0} & = \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\iota_0}), \\ \mathbf{X_1} & = \mathbf{0}_{T\times T}, \\ \mathbf{X_2} & = \mathbf{H}_\phi, \\ \mathbf{z_0} & = \tau_0, \\ \widetilde{\mathbf{y}} & = \mathbf{H}_\phi\mathbf{y}, \\ \widetilde{\boldsymbol{\zeta}} & = \mathbf{0}_{T\times 1}, \\ \widetilde{\boldsymbol{\eta}} & = \mathbf{H} \setminus \boldsymbol{\eta}. \end{cases}$$

•*MNZ-MSOE(UR)*

Recall from Section 2.1 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:

$$\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota_0}\tau_0 + \boldsymbol{\eta}^* + \kappa_\tau\varepsilon + \mathbf{Hc},\\
\mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi\boldsymbol{\iota_0}\tau_0 + \mathbf{H}_\phi\boldsymbol{\eta}^* + (\kappa_\tau\mathbf{H}_\phi + \mathbf{H})\boldsymbol{\varepsilon},\\
\mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}) &= \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\iota_0})\tau_0 + \mathbf{H}_\phi(\mathbf{A} \setminus \boldsymbol{\eta}^*) + \boldsymbol{\varepsilon},\\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{H}_\phi + \mathbf{H},\\
\mathbf{X_0} &= \mathbf{H}_\phi(\mathbf{H} \setminus \boldsymbol{\iota_0}),\\
\mathbf{X_1} &= \mathbf{0}_{T\times T},\\
\mathbf{X_2} &= \mathbf{H}_\phi,\\
\mathbf{z_0} &= \tau_0,\\
\widetilde{\mathbf{y}} &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}),\\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1},\\
\widetilde{\boldsymbol{\eta}} &= \mathbf{A}\setminus\boldsymbol{\eta}^*.
\end{cases}$$

•*MNZ-SSOE*

$$\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \kappa_\tau\boldsymbol{\varepsilon} + \mathbf{Hc},\\
\mathbf{H}_\phi\mathbf{Hy} &= \mathbf{H}_\phi\boldsymbol{\iota_0}\tau_0 + (\kappa_\tau\mathbf{H}_\phi + \mathbf{H})\,\boldsymbol{\varepsilon},\\
\mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}) &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0})\tau_0 + \boldsymbol{\varepsilon},\\
\widetilde{\mathbf{y}} &= \mathbf{X_0}\mathbf{z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}$$

$$\text{s.t.} \begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{H}_\phi + \mathbf{H},\\
\mathbf{X_0} &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}),\\
\mathbf{X_1} &= \mathbf{0}_{T\times T},\\
\mathbf{X_2} &= \mathbf{0}_{T\times T},\\
\mathbf{z_0} &= \tau_0,\\
\widetilde{\mathbf{y}} &= \mathbf{H}_\phi\mathbf{H}(\mathbf{A} \setminus \mathbf{y}),\\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1},\\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}$$

•*Local Level-MSOE*

$$
\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota_0}\tau_0 + \boldsymbol{\eta} + \mathbf{H}\boldsymbol{\varepsilon}, \\
\mathbf{y} &= \mathbf{H} \setminus \boldsymbol{\iota_0}\tau_0 + \mathbf{H}\backslash\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \\
\widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}
$$

$$
\text{s.t.} \begin{cases}
\mathbf{A} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_0} &= \mathbf{H} \setminus \boldsymbol{\iota_0}, \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{I}_{(T\times T)}, \\
\mathbf{z_0} &= \tau_0, \\
\widetilde{\mathbf{y}} &= \mathbf{y}, \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{H}\backslash\boldsymbol{\eta}.
\end{cases}
$$

•*Local Level-SSOE*

$$
\begin{aligned}
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \kappa_\tau\boldsymbol{\varepsilon} + \mathbf{H}\boldsymbol{\varepsilon}, \\
\mathbf{Hy} &= \boldsymbol{\iota}_0\tau_0 + \left(\kappa_\tau\mathbf{I}_{(T\times T)} + \mathbf{H}\right)\boldsymbol{\varepsilon}, \\
\mathbf{H}(\mathbf{A} \setminus \mathbf{y}) &= \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0})\tau_0 + \boldsymbol{\varepsilon}, \\
\widetilde{\mathbf{y}} &= \mathbf{X_0 z_0} + \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} + \mathbf{X_2}\widetilde{\boldsymbol{\eta}} + \boldsymbol{\varepsilon},
\end{aligned}
$$

$$
\text{s.t.} \begin{cases}
\mathbf{A} &= \kappa_\tau\mathbf{I}_{(T\times T)} + \mathbf{H}, \\
\mathbf{X_0} &= \mathbf{H}(\mathbf{A} \setminus \boldsymbol{\iota_0}), \\
\mathbf{X_1} &= \mathbf{0}_{T\times T}, \\
\mathbf{X_2} &= \mathbf{0}_{T\times T}, \\
\mathbf{z_0} &= \tau_0, \\
\widetilde{\mathbf{y}} &= \mathbf{H}(\mathbf{A} \setminus \mathbf{y}), \\
\widetilde{\boldsymbol{\zeta}} &= \mathbf{0}_{T\times 1}, \\
\widetilde{\boldsymbol{\eta}} &= \mathbf{0}_{T\times 1}.
\end{cases}
$$

References

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, 1984.

L. Bauwens, G. Koop, D. Korobilis, and J. V. Rombouts. The Contribution of Structural Break Models to Forecasting Macroeconomic Series. *Journal of Applied Econometrics*, 2014.

S. Beveridge and C. R. Nelson. 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):151–174, 1981.

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

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

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

E. Castelnuovo, L. Greco, and D. Raggi. Policy rules, Regime Switches, and Trend Inflation: An Empirical Investigation for the United States. *Macroeconomic Dynamics*, 18(4):920–942, 2014.

J. C. Chan. Moving Average Stochastic Volatility Models with Application to Inflation Forecast. *Journal of Econometrics*, 176(2):162–172, 2013.

J. C. Chan. The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling. *Journal of Business & Economic Statistics*, 35(1), 2015.

J. C. Chan and I. Jeliazkov. Efficient Simulation and Integrated Likelihood Estimation in State Space Models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1 (1-2):101–120, 2009.

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

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

S. Chib. Markov Chain Monte Carlo Methods: Computation and Inference. *Handbook of Econometrics*, 5:3569–3649, 2001.

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

P. K. Clark. The Cyclical Component of US Economic Activity. *The Quarterly Journal of Economics*, 102(4):797–814, 1987.

T. E. Clark and T. Doh. Evaluating Alternative Models of Trend Inflation. *International Journal of Forecasting*, 30(3):426–448, 2014.

T. E. Clark and F. Ravazzolo. Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, 30(4):551–575, 2014.

J. H. Cochrane. How Big is the Random Walk in GNP? *The Journal of Political Economy*, 96(5): 893–920, 1988.

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

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

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

F. X. Diebold and R. S. Mariano. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144, 1995.

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

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

J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Number 38. Oxford University Press, 2012.

C. S. Forbes, R. D. Snyder, and R. G. Shami. Bayesian Exponential Smoothing. Unpublished Manuscript, Monash University, Department of Econometrics and Business Statistics. 2000.

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

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

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

D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press, 2006.

C. Garnier, E. Mertens, and E. Nelson. Trend Inflation in Advanced Economies. *International Journal of Central Banking*, 11(S1):65–136, 2015.

J. Geweke. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Bernardo, J., Berger, J., Dawid, A. and Smith, A. (eds.). *Bayesian Statistics 4*, pages 641–649, 1992.

J. Geweke and G. Amisano. Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns. *Journal of Applied Econometrics*, 26(1):1–29, 2011.

J. Geweke and C. Whiteman. Bayesian Forecasting. *Handbook of Economic Forecasting*, 1:3–80, 2006.

G. H. Golub and C. F. Van Loan. *Matrix Computations*, volume 3. The Johns Hopkins University Press, Baltimore, 1983.

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

C. W. J. Granger and P. Newbold. *Forecasting Economic Time Series*. Academic Press, 1986.

A. Harvey. Forecasting with Unobserved Components Time Series Models. *Handbook of economic forecasting*, 1:327–412, 2006.

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

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

A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1989.

A. C. Harvey and A. Jaeger. Detrending, Stylized Facts and the Business Cycle. *Journal of Applied Econometrics*, 8:231–231, 1993.

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

K. H. Kang, C.-J. Kim, and J. Morley. Changes in US Inflation Persistence. *Studies in Nonlinear Dynamics & Econometrics*, 13(4), 2009.

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

G. Koop. *Bayesian Econometrics*. Wiley, 2003.

G. Koop, D. J. Poirier, and J. L. Tobias. *Bayesian Econometric Methods*, volume 7. Cambridge University Press, 2007.

S. Luo and R. Startz. Is it One Break or Ongoing Permanent Shocks that Explains US Real GDP? *Journal of Monetary Economics*, 66:155–163, 2014.

M. Marcellino, J. H. Stock, and M. W. Watson. A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics*, 135(1): 499–526, 2006.

W. J. McCausland, S. Miller, and D. Pelletier. Simulation Smoothing For State-Space Models: A Computational Efficiency Analysis. *Computational Statistics & Data Analysis*, 55(1):199–212, 2011.

J. C. Morley, C. R. Nelson, and E. Zivot. Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP so Different? *Review of Economics and Statistics*, 85 (2):235–243, 2003.

K. H. Oh and E. Zivot. The Clark Model with Correlated Components. Unpublished Manuscript, Department of Economics, University of Washington. 2006.

K. H. Oh, E. Zivot, and D. Creal. The Relationship Between the Beveridge-Nelson Decomposition and Other Permanent-Transitory Decompositions that Are Popular in Economics. *Journal of Econometrics*, 146(2):207–219, 2008.

J. K. Ord, A. Koehler, and R. D. Snyder. Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, 92(440):1621–1629, 1997.

J. K. Ord, R. D. Snyder, A. B. Koehler, R. J. Hyndman, and M. Leeds. 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.

O. Papaspiliopoulos, G. Roberts, and M. Söld. Non-Centered Parameterisations for Hierarchical Models and Data Augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, pages 307–326. Oxford University Press, 2003.

P. Perron and T. Wada. Let's Take a Break: Trends and Cycles in US Real GDP. *Journal of Monetary Economics*, 56(6):749–765, 2009.

D. S. G. Pollock, R. C. Green, and T. Nguyen. *Handbook of Time Series Analysis, Signal Processing, and Dynamics*. Academic Press, 1999.

T. Proietti. Trend-cycle Decompositions with Correlated Components. *Econometric Reviews*, 25 (1):61–84, 2006.

F. Smets and R. Wouters. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606, 2007.

R. Snyder. Recursive Estimation of Dynamic Linear Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(2):272–276, 1985.

R. D. Snyder, J. K. Ord, and A. B. Koehler. Prediction Intervals for ARIMA Models. *Journal of Business & Economic Statistics*, 19(2):217–225, 2001.

A. Stella and J. Stock. A State-Dependent Model for Inflation Forecasting. Mimeo. 2013.

J. H. Stock and M. W. Watson. Why Has US Inflation Become Harder to Forecast? *Journal of Money, Credit and banking*, 39(s1):3–33, 2007.

J. H. Stock and M. W. Watson. Core Inflation and Trend Inflation. *Review of Economics and Statistics*, 93(4):770–784, 2016.

T. Teräsvirta. The Invertibility of Sums of Discrete MA and ARMA Processes. *Scandinavian Journal of Statistics*, 4(4):165–170, 1977.

L. Tierney. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4): 1701–1728, 1994.

M. W. Watson. Univariate Detrending Methods with Stochastic Trends. *Journal of Monetary Economics*, 18(1):49–75, 1986.

P. R. Winters. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, 6(3):324–342, 1960.

M. Woodford. How Important is Money in the Conduct of Monetary Policy? *Journal of Money, Credit and Banking*, 40(8):1561–1598, 2007.

V. Zarnowitz and A. Ozyildirim. Time Series Decomposition and Measurement of Business Cycles, Trends and Growth Cycles. *Journal of Monetary Economics*, 53(7):1717–1739, 2006.