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

By

Luis Uzeda

ANU Working Papers in Economics and Econometrics
632

March 2016 JEL: C11, C15, C51, C53

ISBN: 0 86831 632 6

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

Luis Uzeda*
Research School of Economics
Australian National University

March 2, 2016

Abstract

Implications to signal extraction that arise from specifying unobserved components (UC) models with correlated or orthogonal innovations have been well-investigated. In contrast, an analogous statement for forecasting evaluation cannot be made. This paper attempts to fill this gap in light of the recent resurgence of studies adopting UC models for forecasting purposes. In particular, four correlation structures are entertained: orthogonal, correlated, perfectly correlated innovations as well as a novel approach which combines features from two contrasting cases, namely, orthogonal and perfectly correlated innovations. Parameter space restrictions associated with different correlation structures and their connection with forecasting are discussed within a Bayesian framework. Introducing perfectly correlated innovations, however, reduces the covariance matrix rank. To accommodate that, 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.

Keywords: Bayesian, Markov Chain Monte Carlo, State Space, Unobserved Components Models, ARIMA, Reduced Rank, Precision, Forecasting

JEL classification codes: C11, C15, C51, C53

*I would like to thank Rodney Strachan, Joshua Chan, James Morley, Jan Jacobs, Alfred Haug, Tomasz Wozniak and Mohamad Khaled for helpful comments and suggestions. I also thank seminar and conference participants at the Australian National University, the 2^{nd} Workshop of the Australasian Macroeconomic Society, the 2^{nd} Continuing Education in Macroeconometrics Workshop at the University of Tasmania and the 28^{th} PhD in Economics and Business Conference at the University of Queensland.

1 Introduction

Unobserved components (UC) models provide a flexible and yet parsimonious framework which has been widely employed in empirical macroeconomics over the years.[1] When estimating such models, however, one is typically confronted with the issue of formulating the correlation structure between innovations driving different components (states). Depending on the subject matter, economic theory, statistical properties or a combination of both can be used to provide guidance on suitable modelling strategies of the covariance matrix. In particular, there is a wealth of studies discussing different approaches to model correlation and their effects to extracted components (i.e. implications to 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) and Oh et al. (2008) study different correlation structures to reconcile discrepancies in business cycle measures generated by UC models and the ones based on the Beveridge-Nelson decomposition (see Beveridge and Nelson (1981)). Similarly, Dungey et al. (2015) adopt a correlated innovations UC model framework to suggest identification strategies of permanent and transitory shocks to trend and cyclical components of US real GDP.

The brief description above is intended to highlight that a myriad of 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 primary goal of this paper is to fill this gap. Moreover, the recent resurgence of papers adopting UC models for inflation forecasting purposes as in Stock and Watson (2007), Chan, Koop and Potter (2013), Stella and Stock (2013), Garnier, Mertens and Nelson (2013), Chan (2013) and Clark and Doh (2014) strengthens the case for out-of-sample evaluation of the correlation structure within such models.

Notably, modern approaches to forecast using UC models, as in the studies mentioned above, typically exhibit three features: (1) using Bayesian, or more precisely, Markov Chain Monte Carlo (MCMC) techniques to conduct estimation; (2) orthogonal innovations; (3) introducing stochastic volatility a la, e.g., Kim et al. (1998) to model changes in the conditional variance of innovations over time. In this paper we keep point 1, extend point 2 and leave point 3 for future research. To be clear, in our empirical exercise, we decide to leave out stochastic volatility not because

[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.

we think it is unimportant for forecasting, but in order to make the out-of-sample assessment of each correlation structure as free as possible from other modeling features that can also impart changes to forecasting performance. In this sense, once we are confident about potential forecasting benefits of UC models which deviate from the popular orthogonal innovations framework, a natural extension to our study is to incorporate stochastic volatility and conduct the forecasting exercise to a broader range of models.

It is important to recognize that when generating forecasts which incorporate parameter uncertainty, as in the case of Bayesian estimation, the correlation structure has a direct connection with the forecasting function (i.e., the predictive density) which does not arise naturally if adopting other approaches such as maximum likelihood based forecasts. More precisely, within a Bayesian setting, the amount of uncertainty a predictive density can account for is commensurate with the dimension of the parameter space over which an UC model is defined. As a result, depending on how one models correlation between innovations, such parameter space can be more or less restricted and, by the same token, point and density forecasts can be affected. In particular, the usual orthogonal innovations approach often imposes strong parameter space restrictions when compared to their correlated counterparts, as pointed out in, e.g, Harvey and Koopman (2000), Morley et al. (2003), Ord et al. (2005) and Oh et al. (2008). Obviously, whether a more or less restricted parameter space is desirable for forecasting performance is an open empirical question which we address in this paper.

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

To evaluate forecast performance across different correlation structures, a substantive forecasting exercise accounting for eleven UC models is conducted. In keeping with other studies using UC models for forecasting purposes (e.g. Stock and Watson (2007), Chan (2013), Chan (2015),

Clark and Doh (2014) and Garnier et al. (2015)), we use several quarterly inflation measures in our empirical application. Forecasting performance is assessed in terms of point and density forecast accuracy. For the former, we report root mean square forecast error (RMSFE) results while log predictive likelihood results are reported for density forecast performance. 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. All in all, we find that the choice of correlation structure between state variables has appreciable implications to forecasting performance. Allowing for correlation between innovations lead to statistically significant improvements in both point and density forecast estimates at various forecasting horizons relative to orthogonal state variants. 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.

Importantly, allowing for perfectly correlated innovations (as in the SSOE and RSOE cases) produces covariance matrices with reduced rank. To accommodate degeneracies rank reduction can engender to MCMC estimation we develop an MCMC sampler which builds upon properties of Toeplitz matrices and extend previous work on precision-based algorithms for state space models in Chan and Jeliazkov (2009). As shown in McCausland et al. (2011) precision-based samplers are more efficient than the traditional Kalman filter-based approach for state simulation. More precisely, we propose a novel disturbance smoothing algorithm which adds to the (Kalman-filter based) existing ones of DeJong and Shephard (1995) and Durbin and Koopman (2002).

To summarize, four main contributions emerge from our study: (1) a comprehensive discussion of the forecasting implications that stem from specifying UC models with different correlation structures within a Bayesian framework; (2) a new class of UC models that bridges the existing MSOE and SSOE schemes (i.e. RSOE UC models); (3) a novel algorithm for fast disturbance smoothing based on precision samplers; (4) a substantial forecasting application incorporating various and widely used specifications of UC models. The remainder 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 which exhibit a covariance matrix with reduced rank due to two or more states being perfectly correlated. Section 5 develops an efficient and general posterior simulator to estimate UC models with both full and reduced rank covariance matrices. Out-of-sample 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 first difference in log price level measures while I(2)-UC models address inflation dynamics by directly modelling movements in the log price level (e.g. log CPI and log real GDP implicit price deflator).[3] The second approach leads to a larger number of innovations and, consequently, alternative correlation structures relative to I(1)-UC models are explored.[4]

2.1 I(1)-UC models

Now, 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 as *trend inflation*, which accords well with the Beveridge-Nelson characterization of what consists long-run dynamics in macroeconomic aggregates (see Beveridge and Nelson (1981)). Transitory deviations about τ_t are captured by an ergodic autoregressive process, c_t. In keeping with previous studies (e.g. Kang et al. (2009) and Garnier et al. (2015) such transient dynamics provide a measure for the *inflation gap*. To allow for persistent movements in c_t, a p^{th}-order autoregressive lag polynomial, $\phi(L) = (1 - \phi_1 L^1 - \phi_2 L^2 - ... - \phi_p L^p)$, is introduced with roots of $\phi(x) = 0$ lying outside the complex 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.

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

local level, model. Such class of model has been adopted recently by numerous inflation forecasting studies (e.g. Stock and Watson (2007), Chan (2013), Clark and Doh (2014)). In contrast, when c_t is an AR(2) process (as in, e.g., Kang et al. (2009) and Garnier et al. (2015)) persistence in c_t might reflect, for example, the role of nominal price rigidities in slowing down price adjustments about τ_t, as postulated by New Keneysian macroeconomic models (e.g. Christiano et al. (2005) and Smets and Wouters (2007)). In addition, Morley et al. (2003) show that specifying c_t as an AR(2) process enables just-identification of $\rho_{\varepsilon\eta}$.[5] As such, $\rho_{\varepsilon\eta}$ can be inferred using sample information rather than 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 such parameter more difficult.

A common strategy to address identification of $\rho_{\varepsilon\eta}$ is to treat ε_t and η_t as being orthogonal, i.e., $\rho_{\varepsilon\eta} = 0$ (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) suggest a different approach whereby state innovations are perfectly correlated, i.e., $\rho_{\varepsilon\eta} = \pm 1$. Such contrasting identification strategies, however, affect the construction of predictive densities. This issue is investigated more carefully in Section 3 and constitutes the main motivation to our empirical exercise.

To summarize, below we present the three structures entertained in this paper to model the covariance matrix, Ω, associated with I(1)-UC models:

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

The first one denotes the unrestricted case where $\rho_{\eta\varepsilon}$ is estimated. In accordance with the discussion above, such covariance structure only applies to the case where c_t is set as an AR(2) process. The second and third structures describe the perfectly correlated and uncorrelated innovations cases, respectively. Combining all these with $p = 0$ and $p = 2$ gives rise to five I(1)-UC models:

- **Local Level-SSOE**: $\tau_t = \tau_{t-1} + \kappa_\tau\varepsilon_t$, $c_t = \varepsilon_t$,
- **Local Level-MSOE**: $\tau_t = \tau_{t-1} + \eta_t$, $c_t = \varepsilon_t$ $s.t.$ $\mathbb{C}\text{ov}(\varepsilon_t, \eta_t) = 0$,
- **MNZ-SSOE**: $\tau_t = \tau_{t-1} + \kappa_\tau\varepsilon_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$,
- **MNZ-MSOE(UR)**: $\tau_t = \tau_{t-1} + \eta_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$ $s.t.$ $\mathbb{C}\text{ov}(\varepsilon_t, \eta_t) = \sigma_{\eta\varepsilon} \neq 0$,
- **MNZ-MSOE**: $\tau_t = \tau_{t-1} + \eta_t$, $(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_t$ $s.t.$ $\mathbb{C}\text{ov}(\varepsilon_t, \eta_t) = 0$,

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

where MNZ is used here as short notation for Morley, Nelson and Zivot to describe I(1)-UC models with the inflation gap modelled as an AR(2) process. Since such models can accommodate two variants of MSOE schemes, namely, when $\rho_{\varepsilon\eta} = 0$ and $\rho_{\varepsilon\eta}$ is unrestricted, we refer to each case as MNZ-MSOE and MNZ-MSOE(UR), respectively. Also, note that when innovations are allowed to be correlated (i.e. SSOE and MSOE(UR)), I(1)-UC models exhibit a loading parameter, κ_τ. Such parameter governs the correlation sign between states as well as the magnitude of the effect the common innovation, ε_t, has on τ_t. In particular, when $\rho_{\varepsilon\eta}$ is unrestricted we follow Luo and Startz (2014) and specify innovations to τ_t as $\eta_t = \eta_t^* + \kappa_\tau \varepsilon_t$, such that $\eta_t^* \sim \mathcal{N}(0, \sigma_\eta^2)$ and $\mathbb{C}\text{ov}(\varepsilon_t, \eta_t^*) = 0$. 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 which grow over time. As a result, instead of inflation, y_t now denotes different log price level measures and τ_t reflects latent movements in *trend price level* rather than trend inflation. Nevertheless, I(2)-UC models can also be perceived as models for inflation. In fact, note that by taking first differences of y_t and collecting terms accordingly allows us to reexpress the system above as

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

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

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

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

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

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

It should be noted that, if desired, an unrestricted version of Ω for I(2)-UC models could also be estimated. In particular, Oh and Zivot (2006) show that all correlation parameters in (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 simplicity and parsimony 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:

- **Local Linear Trend-SSOE**: $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \kappa_\mu \varepsilon_t$, $c_t = \varepsilon_t$,

- **Local Linear Trend-RSOE**: $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $c_t = \varepsilon_t$,
  $s.t\ \mathbb{C}\text{ov}(\varepsilon_t, \zeta_t) = 0$,

- **Local Linear Trend-MSOE**: $\tau_t = \mu_t + \tau_t + \eta_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $c_t = \varepsilon_t$,
  $s.t\ \mathbb{C}\text{ov}(\varepsilon_t, \eta_t) = \mathbb{C}\text{ov}(\varepsilon_t, \zeta_t) = \mathbb{C}\text{ov}(\eta_t, \zeta_t) = 0$,

[7] For inflation, one possible motivation for such identification scheme is to think of changes in trend inflation as mainly reflecting idiosyncratic systematic changes in the conduct of monetary policy (e.g. Woodford (2008)). Here, broadly captured by ζ_t. On the other hand, ε_t could be perceived as encompassing non-monetary policy factors underlying transitory inflation dynamics. Such transient movements could reflect shifts that affect the (observed or trend) price level but not trend inflation (a plausible scenario when inflation expectations are well-anchored). As an example of such one-off price shifters, one could think of one-off changes in the price level typically observed after changes in taxation (e.g. introduction of value added taxes) and (or) changes in energy and oil prices.

Table 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.

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

- **CLARK-RSOE**: $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $(1 - \phi_1 L - \phi_2 L^2) c_t = \varepsilon_t$,
  $s.t\ \mathbb{C}\text{ov}(\varepsilon_t, \zeta_t) = 0$,

- **CLARK-MSOE**: $\tau_t = \mu_t + \tau_t + \kappa_\tau \varepsilon_t$, $\mu_t = \mu_{t-1} + \zeta_t$, $(1 - \phi_1 L - \phi_2 L^2) c_t = \varepsilon_t$,
  $s.t\ \mathbb{C}\text{ov}(\varepsilon_t, \eta_t) = \mathbb{C}\text{ov}(\varepsilon_t, \zeta_t) = \mathbb{C}\text{ov}(\eta_t, \zeta_t) = 0$,

where the new parameter κ_μ in the SSOE variants 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.

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 $\boldsymbol{\theta}^{UC}$. A density forecast of y_{t+k} is given as

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

follows:[9]

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

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 an UC model can do just that: alter $\boldsymbol{\Theta}^{UC}$.

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

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

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:

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

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

[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)).

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 to the parameter space of a reduced form ARIMA model, $\boldsymbol{\Theta}^{ARIMA}$. To be precise, since $\boldsymbol{\Theta}^{ARIMA}$ is unique, 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 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 generates substantial differences to the support of the predictive likelihood, $f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{ARIMA})$, as measured in terms of the invertibility region of reduced form MA parameters. In particular, looking at the dotted area which describes the support of $f(y_{t+k}|\mathbf{y}, \boldsymbol{\theta}^{ARIMA})$ for the uncorrelated innovations case (i.e. MSOE), one can note that such region is much more restricted relative to the SSOE and RSOE variants. In other words, for the local linear trend model, orthogonality considerably limits the amount of parameter uncertainty $f(y_{t+k}|\mathbf{y})$ can account for.[11] Obviously, whether or not such constraints are desirable for forecasting is, ultimately, an empirical question. Since RSOE models constitute a new class of models, before turning to estimation, we now state the conditions that ensure that such models yield an invertible ARIMA representation:

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

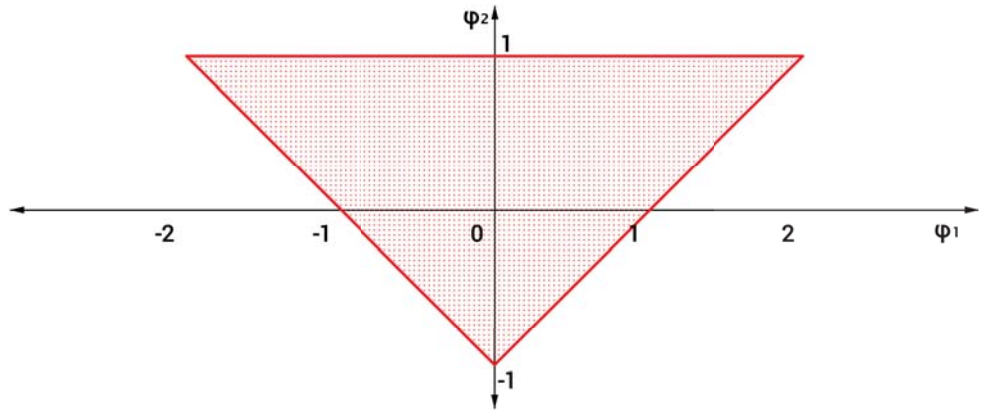
Proof - See appendix A.1.

4 MCMC Inference of UC Models with Reduced Rank Covariance Matrix

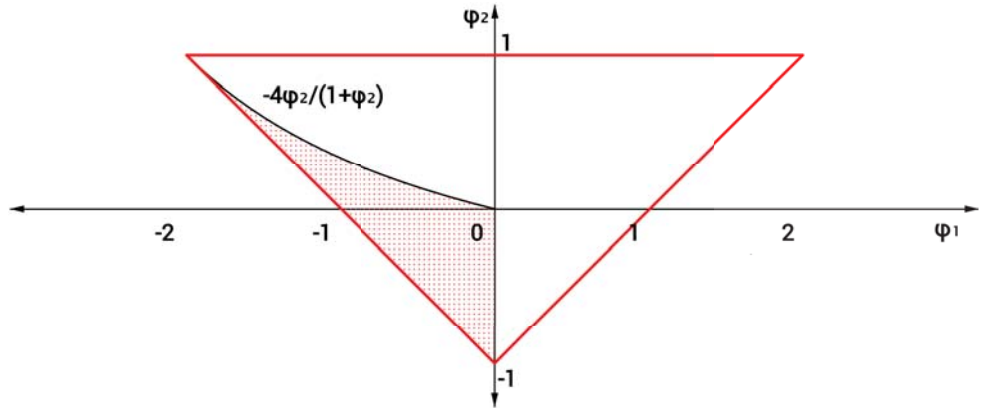
A common approach to estimate UC models (or state space models in general) is to employ MCMC simulation techniques.[12] The usefulness of MCMC sampling in the context of UC models

[11] Derivation of the admissible regions shown in Figure 1 is deferred to a technical appendix.
[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.

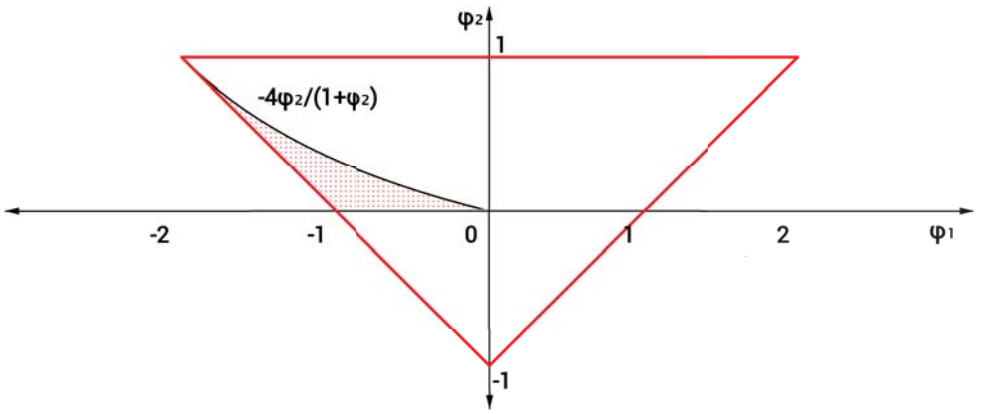
Figure 1: Parameter space restrictions over the invertibility region of reduced-form MA parameters following different correlation structures for the local linear trend model. Dotted area denotes the admissible (or non-constrained) region for each correlation structure.



Local Linear Trend-SSOE



Local Linear Trend-RSOE



Local Linear Trend-MSOE

stems from the modular nature of such type algorithm which transforms 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 which makes direct application of standard MCMC methods problematic. 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.[14] 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, given the data, $\mathbf{y} = (y_1, ..., y_T)'$, and some initial conditions for the states, τ_0 and μ_0, an MCMC algorithm for the local linear trend model can be described as a two-step sampling scheme which involves sequentially drawing from the following conditional posteriors:[15]

$$1. f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}, \tau_0, \mu_0),$$

$$2. f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \tau_0, \mu_0),$$

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

As it is well-known, MCMC simulation only requires evaluation of the *kernel* associated with each conditional posteriors above. Thus, to understand how MCMC sampling can be affected by rank reduction of Ω is is useful to address the implications to the kernel of $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}, \tau_0, \mu_0)$ and $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \tau_0, \mu_0)$. 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* $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}, \tau_0, \mu_0)$ *and* $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \tau_0, \mu_0)$ *exhibits a rank-deficient covariance matrix.*

Proof - See appendix A.2.

[13] Forbes et al. (2000) and Snyder et al. (2001) address Bayesian estimation of state space models with a SSOE representation. Their approach, however, does not encompass MCMC estimation. As it is well known, MCMC estimation allows one to work with a wide range of priors. Moreover, once an MCMC algorithm for SSOE and RSOE models is developed, future work could extend such an algorithm to incorporate nonlinearities which accord well with MCMC sampling, such as stochastic volatility.

[14] The discussion above also applies to all the other UC models in Section 2

[15] For simplicity, at this stage, we assume initial conditions are known and fixed at predetermined values. We discuss initialization of state variables in Section 5

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 the density degeneracies (i.e. probability distributions with zero variance) which occur when allowing for perfectly correlated states. In such cases, simulation smoothing (i.e. sampling from $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}, \tau_0, \mu_0)$) can still be carried out using standard Forward-Filtering-Backward Smoothing (FFBS) algorithms as in Frühwirth-Schnatter (1994), Carter and Kohn (1994), De Jong and Shephard (1995) or Durbin and Koopman (2002). In particular, as pointed out in, e.g., Harvey and Koopman (2000) and Casals et al. (2015) degeneracies which stem from perfectly correlated states can be handled within the Kalman filter by setting the variance filtering step to zero, which makes the backward smoothing step redundant.[16] In contrast, for parameter sampling (i.e. sampling from $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \tau_0, \mu_0)$), 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 novel MCMC sampler which does that by parameterizing UC models in terms of their innovations rather than states.

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 which allows us to recover the states, τ_t, μ_t and c_t by simulating the innovations, η_t, ζ_t and ε_t. In doing so, we construct a general estimation framework which readily accommodates the differences in the covariance rank across models. More precisely, UC models are reparameterized in terms of the disturbances driving the state variables τ_t, μ_t and c_t. Within such a framework, MCMC sampling issues discussed in Section 4 can be easily tackled. As we show below, once we develop an MCMC algorithm for the UC model with the largest number of disturbances (in our case, the CLARK-MSOE model), estimation of all other UC variants can be treated as special cases nested within a general framework.

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

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

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 $\boldsymbol{\Sigma}_i = diag(\sigma_i^2, ..., \sigma_i^2)$ 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

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

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. For simplicity, pre-sample values for the stationary state, c_t, are set to zero.[17]

Next, we derive a disturbance-based parameterization of the model described in (8)-(11). To

[17] Alternatively, 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. Yet another approach could be to initialize c_t using its unconditional first and second moments while setting $\tau_0 = \mu_0 = 0$ and initializing τ_t and μ_t based on diffuse priors for η_1 and ζ_1. A caveat to the latter approach is that, in our setting, some or all of these innovations do not always enter the models (as in the RSOE and SSOE cases). Winters' exponential smoothing method, on the other hand, is easy to implement and can be used for initialization regardless of the number of innovations in each model. Additional approaches to initialize state space models which share some similarities with the framework presented here include De Jong and Chu-Chun-Lin (1994) and Casals and Sotoca (2001).

do so, notice that by pre-multiplying both sides of (8) by $\mathbf{HH}$ gives:

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

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

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

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

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) it is easy to verify that (12d) ensues.[18]

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', so to speak. As a result, the state-space representation in (8)-(11) can be recast as:

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

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

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

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

The usefulness of parameterizing UC models in terms of innovations rather than states (i.e. τ_t, μ_t and c_t) becomes more evident in the case of UC models which contain a covariance matrix with reduced rank. In particular, since rank reduction stems from reducing the number of innovations, the representation in (13)-(15) provides an intuitive way to think of SSOE and RSOE schemes as nested cases within an 'innovations-richer' framework. For example, in the SSOE case one could first set $\boldsymbol{\eta} = \kappa_\tau\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta} = \kappa_\mu\boldsymbol{\varepsilon}$ in (12b) and apply straightforward algebraic manipulations to derive a measurement equation as in (12d), except that now $\mathbf{X_1} = \mathbf{X_2} = \mathbf{0}_{(T\times T)}$ since SSOE models do not require sampling $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\zeta}}$. Naturally, $\widetilde{\mathbf{y}}$ and $\mathbf{X_0}$ would also change accordingly.

In fact, all UC variants entertained in this paper can be accommodated into the framework above for appropriately defined $\widetilde{\mathbf{y}}$, $\widetilde{\boldsymbol{\eta}}$, $\widetilde{\boldsymbol{\zeta}}$, $\mathbf{X_0}$, $\mathbf{X_1}$, $\mathbf{X_2}$ and $\mathbf{z_0}$. Doing so reduces the coding burden

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

typically associated with adapting an MCMC algorithm to various problems.[19] To avoid cluttering the discussion here with algebraic details we defer the derivation and presentation of the exact structures of such matrices and vectors for each model to a technical appendix.

Obtaining posterior draws for the representation in (13)-(15) can, thus, be summarized as a three-step algorithm which requires sequentially sampling from:[20]

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

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

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

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

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

[20] For notational convenience we exclude $\mathbf{z_0}$ as a conditioning factor in conditional posteriors since initial conditions denote deterministic functions of the first observations in $\mathbf{y}$.

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

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

$$\begin{aligned} f(\widetilde{\boldsymbol{\eta}}|\mathbf{y}, \mathbf{z}_{-\tilde{\eta}}, \boldsymbol{\theta}) &\propto \exp\left(-\frac{(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}})'\boldsymbol{\Sigma}_{\varepsilon}^{-1}(\widetilde{\mathbf{y}}_{\tilde{\eta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}}) + \widetilde{\boldsymbol{\eta}}'\mathbf{D}_{\tilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}}}{2}\right) \\ &\propto \exp\left(-\frac{\widetilde{\boldsymbol{\eta}}'\left(\mathbf{X}_2'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)\widetilde{\boldsymbol{\eta}} - 2\widetilde{\mathbf{y}}_{\tilde{\eta}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X_2}\widetilde{\boldsymbol{\eta}}}{2}\right) \\ &= \exp\left(-\frac{\widetilde{\boldsymbol{\eta}}'\overline{\mathbf{D}}_{\tilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}} - 2\bar{\mathbf{d}}_{\tilde{\eta}}'\overline{\mathbf{D}}_{\tilde{\eta}}^{-1}\widetilde{\boldsymbol{\eta}}}{2}\right). \end{aligned}$$

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 $\overline{\mathbf{D}}_{\tilde{\eta}} = \left(\mathbf{X}_2'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)^{-1}$ and $\bar{\mathbf{d}}_{\tilde{\eta}} = \overline{\mathbf{D}}_{\tilde{\eta}}\mathbf{X}_2'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\widetilde{\mathbf{y}}_{\tilde{\eta}}$. 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 $\overline{\mathbf{D}}_{\tilde{\eta}} = \left(\mathbf{X}_2'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)^{-1}$ 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 $\left(\mathbf{X}_2'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X_2} + \mathbf{D}_{\tilde{\eta}}^{-1}\right)^{-1}$ via brute-force inversion (see e.g. Golub and Van Loan (1983) p.156).

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

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

The first step describes the Cholesky decomposition of $\overline{\mathbf{D}}_{\tilde{\eta}}^{-1}$, such that $\overline{\mathbf{D}}_{\tilde{\eta}}^{-1} = \mathbf{C}\mathbf{C}'$. Since $\overline{\mathbf{D}}_{\tilde{\eta}}^{-1}$ is a banded matrix, a Cholesky factorization only involves $\mathcal{O}(T)$ operations (see Golub and Van Loan (1983) p.156). Step 2 requires solving a triangular system by forward substitution (given that $\mathbf{C}$

is a lower triangular matrix) which entails $\mathcal{O}(T)$ operations as well. Step 3 is equivalent to Step 2, except that the solution of the triangular system, $\mathbf{C}'\backslash\mathbf{x}$, is now obtained by backward substitution. It is then straightforward to see that Steps 2 and 3 combined, by definition, yield

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

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 $\widetilde{\boldsymbol{\eta}} \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\widetilde{\eta}}, \overline{\mathbf{D}}_{\widetilde{\eta}}\right)$. 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 $(\widetilde{\boldsymbol{\zeta}}|\mathbf{y}, \mathbf{z}_{-\widetilde{\zeta}}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\bar{\mathbf{d}}_{\widetilde{\zeta}}, \overline{\mathbf{D}}_{\widetilde{\zeta}}\right)$, where $\overline{\mathbf{D}}_{\widetilde{\zeta}} = \left(\mathbf{X}_{\mathbf{1}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X}_{\mathbf{1}} + \mathbf{D}_{\widetilde{\zeta}}^{-1}\right)^{-1}$ and $\bar{\mathbf{d}}_{\widetilde{\zeta}} = \overline{\mathbf{D}}_{\widetilde{\zeta}}\mathbf{X}_{\mathbf{2}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\widetilde{\mathbf{y}}_{\widetilde{\zeta}}$, where $\widetilde{\mathbf{y}}_{\widetilde{\zeta}} = \widetilde{\mathbf{y}} - \mathbf{X}_{\mathbf{0}}\mathbf{z}_{\mathbf{0}} - \mathbf{X}_{\mathbf{2}}\widetilde{\boldsymbol{\eta}}$. Note that $\mathbf{X}_{\mathbf{1}} = \mathbf{H}_{\phi}$, $\mathbf{X}_{\mathbf{1}}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{X}_{\mathbf{1}}$ and $\mathbf{D}_{\widetilde{\zeta}}^{-1} = (\mathbf{H}\mathbf{H})'\boldsymbol{\Sigma}_{\zeta}^{-1}(\mathbf{H}\mathbf{H})$ are all sparse matrices. Therefore, draws of $\widetilde{\boldsymbol{\zeta}} = (\mathbf{H}\mathbf{H})^{-1}\boldsymbol{\zeta}$ can also be quickly obtained following the four $\mathcal{O}(T)$ steps described above. Finally, by setting $\boldsymbol{\zeta} = \mathbf{H}\mathbf{H}\widetilde{\boldsymbol{\zeta}}$ allows one to recover $\boldsymbol{\zeta}$.

5.2 Parameter Sampling

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

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

two-step algorithm:

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

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

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

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

where $\mathcal{IG}$ denotes an inverse-gamma density and $\mathbb{I}_{(\psi \in A_\psi)}$ and $\mathbb{I}_{(\phi \in A_\phi)}$, respectively, represent indicator functions which ensure draws of κ_i and ϕ_i are compatible with an invertible and stationary reduced form ARIMA representation of the UC models in Section 2.[22]

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. In fact, using standard methods (see e.g. Koop (2003)) one can verify that

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

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

Next, draws of the loading and autoregressive parameters require simulating $f(\boldsymbol{\theta_{-\sigma^2}}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$. The latter, however, is not of a known form which one can readily sample from. To circumvent this issue, we introduce a Metropolis-Hastings step to our algorithm. To do so, note first that combining the likelihood function in (16) with a joint prior density $f(\boldsymbol{\theta_{-\sigma^2}})$ yields:

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

where $\widetilde{\mathbf{y}}_* = \widetilde{\mathbf{y}} - \mathbf{X_0 z_0} - \mathbf{X_1}\widetilde{\boldsymbol{\zeta}} - \mathbf{X_2}\widetilde{\boldsymbol{\eta}}$. Notably, depending on the UC model specification, parameters in $\boldsymbol{\theta_{-\sigma^2}}$ show up in different components of the measurement equation in (13). In the CLARK-MSOE model, for example, recall from our previous discussion that $\boldsymbol{\theta_{-\sigma^2}} = \{\phi_1,\ \phi_2\}$, $\mathbf{X_0} = ((\mathbf{HH})^{-1}\mathbf{H}_\phi \mathbf{H}\boldsymbol{\iota}_0 \quad (\mathbf{HH})^{-1}\mathbf{H}_\phi \boldsymbol{\iota}_0)$, and $\mathbf{X_1} = \mathbf{X_2} = \mathbf{H}_\phi$.[23] Nonetheless, regardless of the specification all such matrices are banded. Hence, $\log f(\boldsymbol{\theta_{-\sigma^2}}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma^2})$ can be quickly evaluated

[21] Prior hyperparameters are presented in Section 6.1

[22] See appendix for a discussion on the invertibility and stationary conditions required for the UC models adopted in this paper.

[23] 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 the other UC models employed in this paper.

using sparse routines and the generalized inverse operator (\) discussed in Section 5.1, both implemented in most statistical packages. Then, 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 and the negative Hessian evaluated at the mode of $f(\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2}|\mathbf{y}, \mathbf{z}, \boldsymbol{\sigma}^2)$, denoted as $\hat{\boldsymbol{\theta}}_{-\boldsymbol{\sigma}^2}$ and $\mathbf{G}^{-1}$, respectively.

6 Evaluation

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

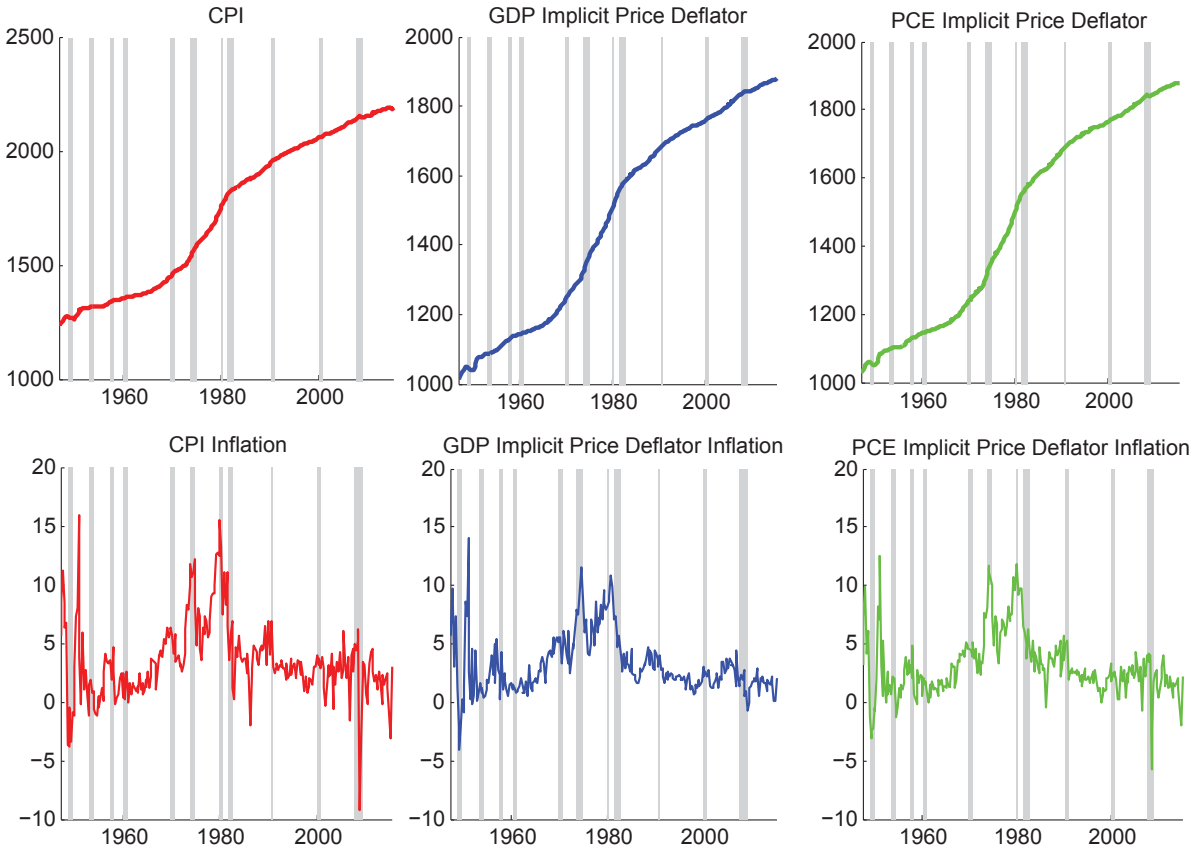
6.1 Data and Priors

Akin to Stock and Watson (2007), our data set consists of quarterly series for CPI and the implicit price deflators for real GDP and personal consumption expenditure (PCE). Quarterly series were constructed by averaging monthly data over the 3 months within the quarter. In keeping with the literature, given a quarterly figure, x_t, we use $y_t = 400\log(x_t)$ to capture price changes at a quarterly annualized basis. Again, remember that for I(1)-UC models y_t denotes the first difference in logs of a price level measure while for I(2)-UC models y_t denotes the actual log-price level measure. Figure 2 shows the data used in our empirical exercise.

Priors are selected to balance three criteria: facilitate comparison across models, being relatively uninformative and follow recommendations from previous studies. For example, in keeping with previous forecasting studies (e.g., Stock and Watson (2007), Chan (2013) and Clark and Doh (2014)) we follow the practice of using inverse-gamma priors for variance parameters with hyperparameters calibrated to reflect reasonably uninformative priors. An exception to that is σ_η^2 for I(2)-UC models. In particular, we follow the recommendation in Zarnowitz and Ozyildirim (2006) who suggest filtering I(2) processes with an UC representation that assign quite small conditional variance of the trend level state (i.e., τ_t).[24] 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 hyperparameters for each model.

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

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



Note: Shaded regions indicate recessions as recorded by the NBER

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	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\phi)}$	–	–
Local Linear Trend-RSOE	$\mathcal{IG}(10,9)$	–	$\mathcal{IG}(10,9)$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	–	–	–
Local Linear Trend-MSOE	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9^{-10^6})$	$\mathcal{IG}(10,9)$	–	–	–	–
CLARK-SSOE	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\psi\in A_\phi)}$
CLARK-RSOE	$\mathcal{IG}(10,9)$	–	$\mathcal{IG}(10,9)$	$\mathcal{N}(0,10)\mathbb{I}_{(\psi\in A_\psi)}$	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$
CLARK-MSOE	$\mathcal{IG}(10,9)$	$\mathcal{IG}(10,9^{-10^6})$	$\mathcal{IG}(10,9)$	–	–	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$	$\mathcal{N}(0,0.01)\mathbb{I}_{(\phi\in A_\phi)}$

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.[25] 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 RMSFE associated with each model. Density forecasts are evaluated in terms of average 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 which, under the null hypothesis, postulates equivalent forecasting accuracy between competing models. To control for serial correlation in forecast errors, akin to Clark and Doh (2014), standard errors of t-statistics are computed using a heteroskedasticity and autocorrelation-consistent (HAC) with pre-whitened quadratic spectral estimator. Such test is applied for both RMSFE and predictive log-score results.[26]

To generate forecasts we conduct a predictive simulation exercise along the lines of Cogley et al. (2005) adapted to the disturbance-based parametrization discussed is 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, akin to the discussion in Section 3 and using standard results for conditional probability, a k-step ahead predictive density can be expressed as:

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

where

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

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,

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

[26] 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.

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, one can readily evaluate $f(\Delta y_{t+k}|\mathbf{y_{1:t}}, \mathcal{F}^{(r)})$ to approximate $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$. In fact, our predictive sampler can be summarized as follows:

At *every r-th* iteration of our MCMC algorithm and given data up to time t we (sequentially):
Step 1: simulate parameters and a $t \times 1$ vector of innovations based on the MCMC algorithm discussed in Section 5.
Step 2: simulate future innovations in $\mathbf{z_{t+1:t+k}}$ from the joint normal density, $f(\mathbf{z_{t+1:t+k}}|\mathbf{y_{1:t}}, \mathbf{z_{1:t}}, \boldsymbol{\theta})$, using variance parameters obtained in Step 1.
Step 3: use parameters and innovations produced in Steps 1 and 2 to generate draws for $\mathbf{\Delta y_{t+1:t+k}}$ from the joint normal density, $\prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta})$.

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

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

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

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

for any given initial conditions), residual estimates of ε_t allow one to recover c_t (since $\phi(L)c_t = \varepsilon_t$), which yields τ_t (since $\tau_t = y_t - c_t$) and returns μ_t (since $\mu_t = \tau_t - (\tau_{t-1} + \eta_t)$).

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

6.3 Results: Point Forecasts

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

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

where y^o_{t+k} denotes the observed value of y_{t+k} that is known at time $t+k$, while t_0 and $T-k$ denote the first and last forecast generated respectively. Also, $\widehat{\mathbb{E}}(y_{t+k}|\mathbf{y_{1:t}})$ (i.e. an estimate of the mean of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$) is constructed by averaging over R draws from $\prod_{s=1}^{k} f(\Delta y_{t+s}|\mathbf{y_{1:t+s-1}}, \mathbf{z_{1:t+k}}, \boldsymbol{\theta})$ generated by the relevant innovations-based forecasting parameterization of an UC model, as discussed above.[27] Next, recall that we set t_0 =1971Q1 and T =2015Q2.

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 which outperform the local level MSOE model. Also, numbers in bold denote the best performing model for a specific forecast horizon.

Overall, results in Table 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

[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.

improves point forecast results at longer horizons for all measures of inflation we investigate. In particular, the local linear trend-RSOE model (i.e. when innovations to y_t and τ_t are perfectly correlated, but μ_t is still orthogonal) emerged as the best model at long run (point) forecasting. Such result is likely to reflect the usefulness of smoother measures of trend inflation (see Figure 3) in capturing long run inflation dynamics, as discussed in, e.g., Chan (2013) and Clark and Doh (2014). Notably, best performing models generate improvements which are statistically significant at the 5% level relative to the orthogonal local level model. Also, I(2)-UC models which allow for some or all innovations to be correlated, such as the local linear trend-SSOE and CLARK-RSOE variants, outperform both the baseline model as well as their orthogonal counterparts at two, three and four year-ahead forecast horizons.

To summarize, results in Table 3 point to two main recommendations: (a) Parsimonious UC models with orthogonal innovations seem appropriate for short-horizon point forecasts; (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.[28]

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 which display equivalent means but differ in terms of their variances. In an RMSFE, predictions from both densities are equivalent. On the other hand, when using forecast metrics which incorporate uncertainty around the prediction location the predictive accuracy between such densities would differ commensurate with their difference in variances.

Now, recall from 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 an UC model accommodates, as measured in terms of the restrictions an UC model specification imposes over its corresponding reduced form ARIMA parameter space. Since changes in the correlation structure alter the parameter space over which $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$ is defined, forecast metrics which incorporate information within the full support of $f(\Delta y_{t+k}|\mathbf{y_{1:t}})$, are, thus, more likely to reflect stronger implications between out-of-sample performance and correlation structure.

[28] Point (b) also suggests that depending on the forecast horizon of interest, one should consider directly modelling the price level (i.e. for long-run forecasts). Combining different correlation structures and price level modeling with features such as stochastic volatility might be an interesting avenue for future research.

Table 3: Relative RMSFEs for US Quarterly Inflation Measures

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

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

A natural candidate to evaluate density forecasts is the sum of log predictive likelihoods:[29]

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

For this metric, if the actual outcome $y^o_{t+k}|\mathbf{y_{1:t}}$ is unlikely under the density forecast, the value of $\log(f(\Delta y_{t+k} = \Delta y^o_{t+k}|\mathbf{y_{1:t}}))$ will be small and vise versa. Therefore, larger values of the sum of log predictive likelihoods indicate superior forecast performance. As before, for easy comparison, we present density forecast results relative to a baseline model given by the local level MSOE model. Specifically, entries in Table 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 reinforces the idea that modeling state correlation unambiguously affects density forecast performance at all horizons. In particular the Local level-SSOE is the best model for 1Q-ahead forecast, while the local linear trend-RSOE emerged as the best model for medium to longer forecasting horizons. As in the point forecast context, improvements for the best performing models are also statistically significant at the 5% level relative to the orthogonal local level baseline.

6.5 Correlation and Measures of Trend Inflation

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

Figure 3 shows that differences in trend inflation between MSOE and SSOE variants are minor. Trend inflation measures based on SSOE models are slightly more erratic than MSOE variants. Such result most likely reflects the fact that latent components in the SSOE case are recovered from an unique source of randomness which encompasses all explained variability in inflation (or the price level). When comparing measures of trend inflation between MNZ-MSOE(UR) (which allows for imperfect correlation between c_t and τ_t) and its SSOE counterpart, differences are virtually imperceptible. One possible explanation for this result is presented in Figure 4, which shows that the implied posterior correlation between τ_t and c_t piles up near the (positive) perfect correlation region.

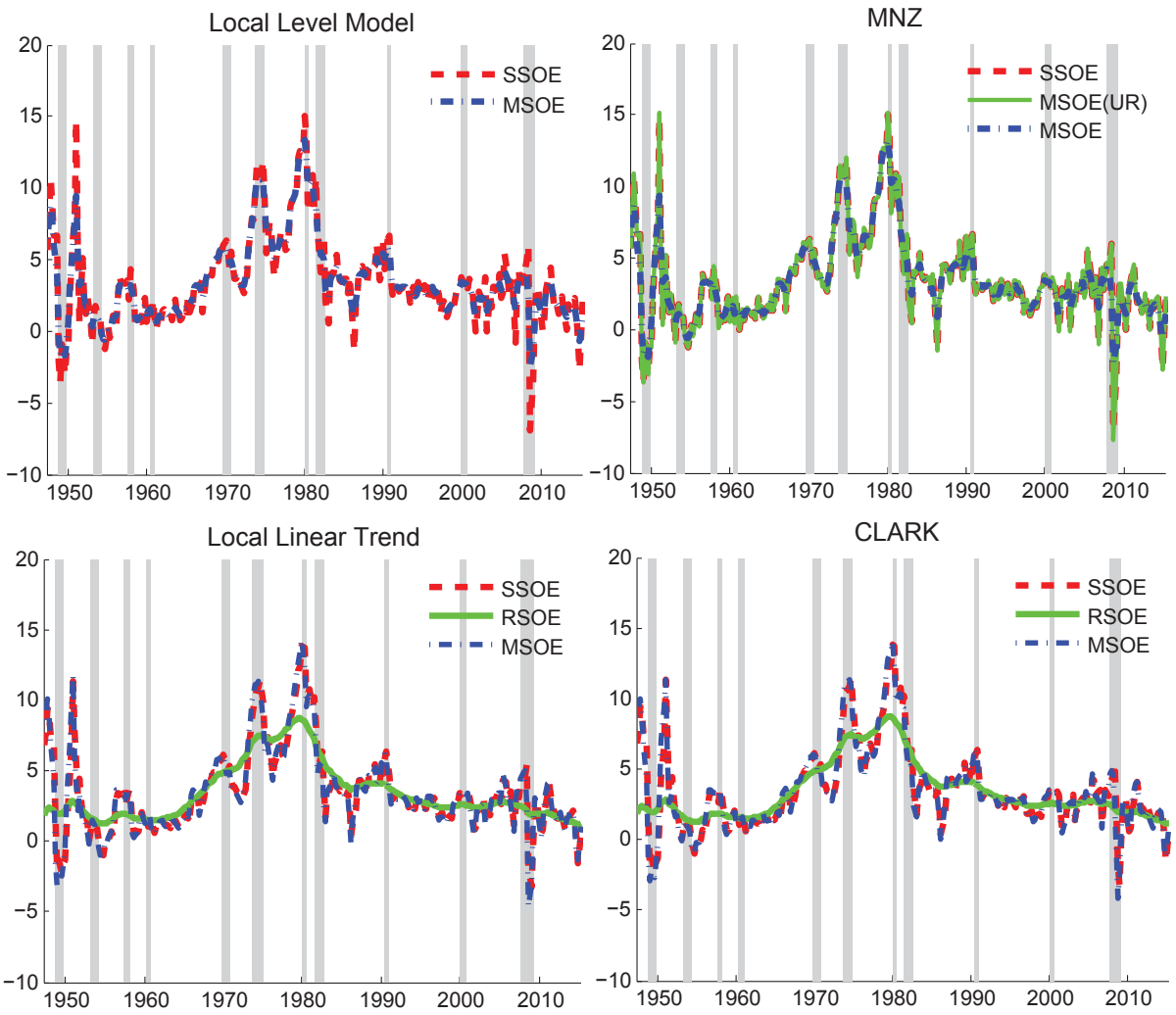
[29] See, e.g., Geweke and Amisano (2011) for a discussion on the predictive likelihood and its usefulness as a model comparison device.

Table 4: Sum of Log Predictive Likelihoods for US Quarterly Inflation Measures

CPI Inflation						
	Forecast Horizon					
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	**31.5***	29.1	26.6	19.1	16.4	17.8
Local Level-MSOE	0.0	0.0	0.0	0.0	0.0	0.0
MNZ-SSOE	7.7	5.1	9.3	9.2	6.1	4.3
MNZ-MSOE(UR)	13.6	16.1	18.7	18.1	20.3	11.1
MNZ-MSOE	-7.3	-10.2	-26.1	-13.1	9.1	5.2
Local Linear Trend-SSOE	14.3	12.1	13.2	10.4	9.7	15.6
Local Linear Trend-RSOE	11.7	13.5	**49.0***	**40.8***	**30.1***	**29.1***
Local Linear Trend-MSOE	-12.4	-10.1	-16.4	-20.4	-28.9	-41.4
CLARK-SSOE	27.1	**40.1***	41.1	39.1	27.4	26.1
CLARK-RSOE	29.1	39.8	24.9	29.6	38.8	39.1
CLARK-MSOE	-15.5	11.2	12.6	19.1	6.1	10.8
GDP-Implicit Price Deflator Inflation						
	Forecast Horizon					
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	**33.1***	24.2	23.1	21.9	19.1	19.8
Local Level-MSOE	0.0	0.0	0.0	0.0	0.0	0.0
MNZ-SSOE	7.4	7.2	8.1	8.3	7.7	5.1
MNZ-MSOE(UR)	14.1	17.2	15.9	18.5	19.3	11.7
MNZ-MSOE	-9.1	-13.6	-25.8	-18.2	3.1	2.0
Local Linear Trend-SSOE	14.3	14.7	11.9	20.1	17.4	13.1
Local Linear Trend-RSOE	10.1	18.9	28.5	**38.9***	**35.0***	**32.7***
Local Linear Trend-MSOE	-9.4	-12.4	-11.7	-16.5	-18.4	-30.8
CLARK-SSOE	29.1	**37.7***	**36.5***	34.1	29.1	24.9
CLARK-RSOE	22.5	29.1	25.4	21.5	30.3	29.5
CLARK-MSOE	-8.9	4.2	6.9	15.7	9.3	11.4
PCE-Implicit Price Deflator Inflation						
	Forecast Horizon					
Model	1Q	2Q	1 Year	2 Years	3 Years	4 Years
Local Level-SSOE	**35.1***	25.3	23.1	29.4	19.1	20.9
Local Level-MSOE	0.0	0.0	0.0	0.0	0.0	0.0
MNZ-SSOE	6.1	6.5	10.9	11.2	8.3	9.1
MNZ-MSOE(UR)	11.3	18.2	15.7	18.7	21.9	20.2
MNZ-MSOE	-8.9	-11.1	-16.5	-17.3	8.2	7.1
Local Linear Trend-SSOE	11.4	19.9	**41.9***	34.1	37.1	33.1
Local Linear Trend-RSOE	19.1	23.1	34.5	**37.9***	**41.4***	**44.6***
Local Linear Trend-MSOE	-10.1	-18.9	-11.3	-12.1	-18.9	-36.4
CLARK-SSOE	17.9	**33.3***	36.1	38.1	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: US Quarterly Measures of Annualized Trend (CPI) Inflation



Note: Shaded regions indicate recessions as recorded by the NBER

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

6.6 Computation Efficiency

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

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

In practice, one alternative to improve MCMC simulation efficiency for UC models is to reparameterize the standard state-based representation given in Section 2.[30] Therefore, albeit UC models are reparameterized in this paper with intent to address rank-reduction issues, a natural question is whether or not the disturbance-based parameterization combined with precision sampling techniques discussed in Section 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:

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

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

Figure 4: Density Estimate for the Correlation Between τ_t and c_t Under MNZ-MSOE(UR) Model

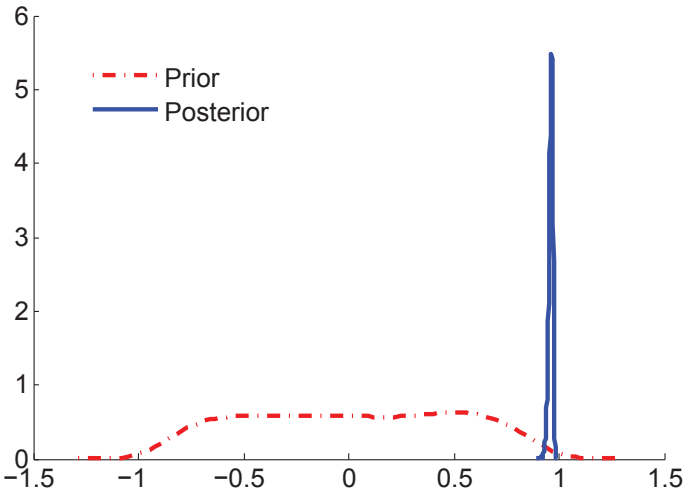
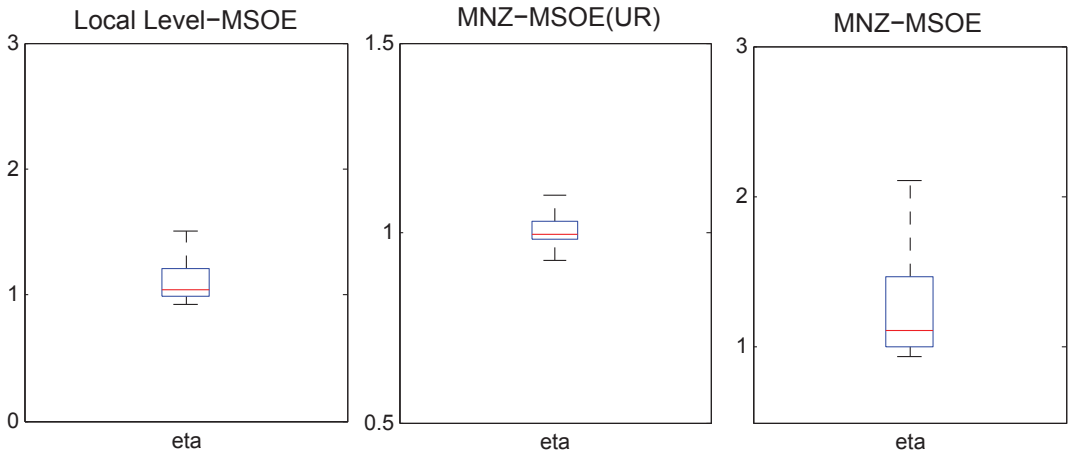


Figure 5: Inefficiency Factors for Disturbance Smoothing (I(1)-UC Models)



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 1. As a rule of thumb, inefficiency factors around 20 based on the metric above are typically interpreted as an indication of fast mixing.[31] 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 present the information on inefficiency factor visually. In particular, the middle line denotes the median inefficiency factor based on a sample T inefficiency factors constructed for each element in $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$. Similarly, lower and upper lines respectively represent the 25 and 75 percentiles, while whiskers extend to the maximum and minimum. All and in all, our results below suggest that the algorithm developed in Section 5 is quite efficient in terms of generating parameters and innovations draws which are not strongly autocorrelated. In addition to inefficiency factors, other metrics of interest to assess MCMC sampling performance are the acceptance ratio on the proposal density in the Metropolis-Hastings step as well as computational speed of our algorithm. Acceptance rates range from 70% to 90% depending on the model, thus, suggesting that the Gaussian proposal well approximates the conditional posterior density of $\boldsymbol{\theta}_{-\boldsymbol{\sigma}^2}|\widetilde{\mathbf{y}}, \mathbf{z}, \boldsymbol{\sigma}^2$. In terms of computational speed, estimation of all models where 10000 draws are sampled for each model takes less than 180 seconds.

[31] 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 6: Inefficiency Factors for Disturbance Smoothing (I(2)-UC Models)

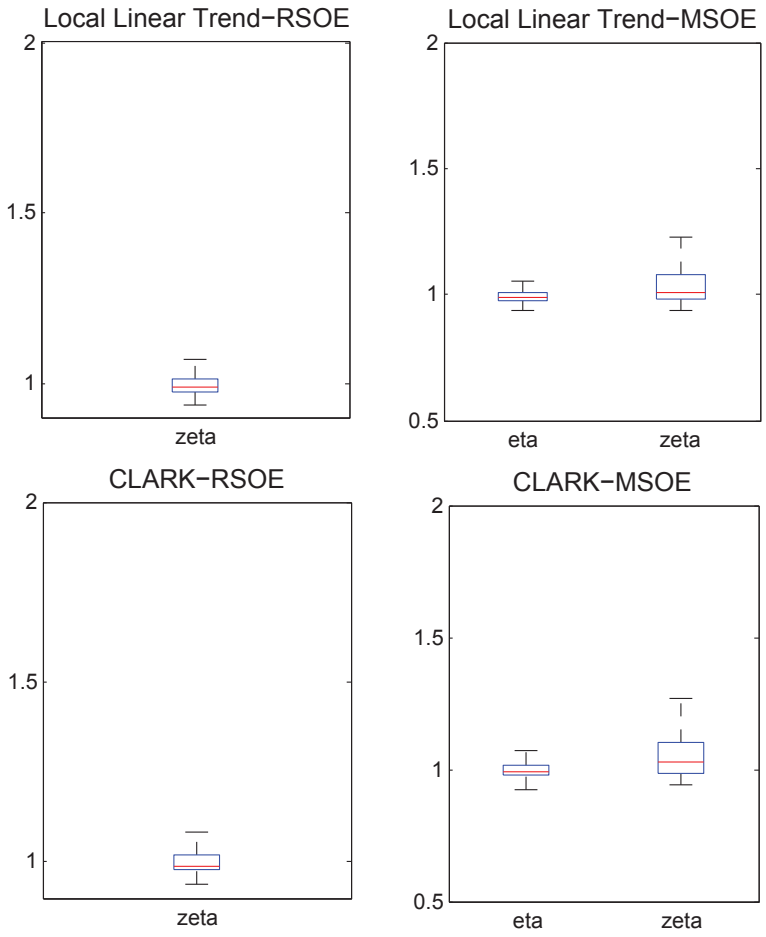


Table 5: Inefficiency Factors for Parameter Sampling

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

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 model. Akin to Stock and Watson (2007), in our empirical application we used inflation measures based on the CPI and the real GDP and PCE price deflators.

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

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

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 a MA(0) (i.e. ζ_t) and a MA(2) (i.e. $\left[\kappa_\tau(1-L) + 1 - 2L + L^2\right]\varepsilon_t$) process. Granger's lemma, thus, ensures that the resulting process will be an MA(q) polynomial such that $q = \max(0, 2) = 2$.

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

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 a MA(3) reduced form representation. To ensure the resulting MA polynomial is invertible, we use use Theorem 1 in Teräsvirta (1977), which states that the sum of two or more (possibly correlated) MA polynomials is invertible if and only if such polynomials do not share common roots of modulus one. The latter is satisfied by noting that the stationarity condition of the AR polynomial precludes $\phi(\pm 1) = 0$. Therefore, as long as c_t is a stationary state, the MA polynomials above, $\phi(L)\zeta_t$ and

$[\kappa_\tau \phi(L)(1-L) + (1-L)^2]\,\varepsilon_t$, 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:

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

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

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 show in Figure 1. To that end, for the MSOE and RSOE variants we take a similar approach as in, e.g., Watson (1986) and Harvey (1989), thereby comparing the autocorrelation function generated by the local linear trend model with its corresponding (unrestricted) reduced form counterpart. Remember, underpinning such comparison is the fact that UC models can also be perceived as structural representations of an ARIMA model. Hence, let $\rho(s)$ denote the autocorrelation function associated with the reduced form ARIMA (0,2,2) in (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 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 which intersect with the $\varphi_1 < 0$ and $\varphi_2 > 0$ quadrant and the the invertibility space (i.e., the triangular region). The resulting region is the admissibility region of MA parameters for the Local Linear Trend-MSOE

model as shown in the bottom panel in Figure 1.

• *Local Linear Trend-RSOE*

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

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

•*Local Level-SSOE*

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

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

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

•*Local Linear Trend-SSOE*

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

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

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

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

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

•*MNZ-SSOE*

Recall that we set $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$. Taking first difference of the measurement equation

for the MNZ-SSOE model and rearranging terms yields:

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

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

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 ARIMA(2,2,3) process above is the same regardless if it is presented in terms of ε_t (i.e. the SSOE representation of the ARIMA model) or u_t. Hence, keeping draws of ϕ_1 and ϕ_2 such that $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$, ensures stationarity of the AR(2) polynomial, $\phi(L)$.

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 which is conducive to quick MCMC estimation. In the interest of brevity, we do not repeat the algebra for the CLARK-MSOE model here. Also, note that the derivation starting point for I(2)-UC models is equation (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, adjusting each UC model according to its correlation structure, as discussed in Section 2, is ensued by straightforward algebraic manipulations.

I(2)-UC Models

•*CLARK-RSOE*

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

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

•*CLARK-SSOE*

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

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

•*Local Linear Trend-MSOE*

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

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

•*Local Linear Trend-RSOE*

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

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

•*Local Linear Trend-SSOE*

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

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

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 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:

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

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

•*MNZ-SSOE*

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

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

•*Local Level-MSOE*

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

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

•*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

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

Luc Bauwens, Gary Koop, Dimitris Korobilis, and Jeroen VK Rombouts. The Contribution of Structural Break Models to Forecasting Macroeconomic Series. *Journal of Applied Econometrics*, 2014.

Stephen Beveridge and Charles 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.

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

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

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

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

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

Joshua CC Chan and Ivan 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.

Chris Chatfield, Anne B Koehler, J Keith Ord, and Ralph 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.

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

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

Peter K Clark. The Cyclical Component of US Economic Activity. *The Quarterly Journal of Economics*, pages 797–814, 1987.

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

Todd E Clark and Francesco Ravazzolo. Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, 2014.

John H Cochrane. How Big is the Random Walk in GNP? *The Journal of Political Economy*, pages 893–920, 1988.

Timothy Cogley, Sergei Morozov, and Thomas 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.

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

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

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 1995.

Mardi Dungey, Jan P.A.M. Jacobs, Jing Tian, and Simon 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, 6 2015.

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

James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*. Number 38. Oxford University Press, 2012.

Catherine S Forbes, Ralph D Snyder, and Roland G Shami. Bayesian Exponential Smoothing. *Unpublished Manuscript, Monash University, Department of Econometrics and Business Statistics*, 2000.

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

Sylvia 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.

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

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

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

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

John Geweke and Gianni Amisano. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26(1):1–29, 2011.

John Geweke and Charles Whiteman. Bayesian forecasting. *Handbook of Economic Forecasting*, 1:3–80, 2006.

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

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

Clive William John Granger and Paul Newbold. *Forecasting Economic Time Series*. Academic Press, 1986.

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

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

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

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

Kyu Ho Kang, Chang-Jin Kim, and James Morley. Changes in US Inflation Persistence. *Studies in Nonlinear Dynamics & Econometrics*, 13(4), 2009.

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

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

Gary Koop, Dale J Poirier, and Justin L Tobias. *Bayesian Econometric Methods*, volume 7. Cambridge University Press, 2007.

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

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

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

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

Kum Hwa Oh and Eric Zivot. The Clark Model with Correlated Components. *Unpublished Manuscript, Department of economics, University of Washington*, 2006.

Kum Hwa Oh, Eric Zivot, and Drew 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 Keith Ord, Ralph D Snyder, Anne B Koehler, Rob J Hyndman, and Mark 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.

John Keith Ord, AB Koehler, and Ralph D Snyder. Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, 92 (440):1621–1629, 1997.

O. Papaspiliopoulos, G.O. 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, USA, 2003.

Pierre Perron and Tatsuma Wada. Let's Take a Break: Trends and Cycles in US Real GDP. *Journal of Monetary Economics*, 56(6):749–765, 2009.

Daniel Stephen Geoffrey Pollock, Richard C Green, and Truong Nguyen. *Handbook of Time Series Analysis, Signal Processing, and Dynamics*. Academic Press, 1999.

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

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

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

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

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

Timo Teräsvirta. The Invertibility of Sums of Discrete MA and ARMA Processes. *Scandinavian Journal of Statistics*, pages 165–170, 1977.

Luke Tierney. Markov Chains for Exploring Posterior Distributions. *the Annals of Statistics*, pages 1701–1728, 1994.

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

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

Michael Woodford. How Important is Money in the Conduct of Monetary Policy? *Journal of Money, Credit and Banking*, 40(8):1561–1598, 2008.

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