

# CAMA

Centre for Applied Macroeconomic Analysis

---

## Bayesian State Space Models in Macroeconometrics

---

CAMA Working Paper 90/2020  
October 2020

**Joshua C.C. Chan**

Purdue University  
University of Technology Sydney  
Centre for Applied Macroeconomic Analysis, ANU

**Rodney W. Strachan**

University of Queensland  
Centre for Applied Macroeconomic Analysis, ANU

### Abstract

State space models play an important role in macroeconometric analysis and the Bayesian approach has been shown to have many advantages. This paper outlines recent developments in state space modelling applied to macroeconomics using Bayesian methods. We outline the directions of recent research, specifically the problems being addressed and the solutions proposed. After presenting a general form for the linear Gaussian model, we discuss the interpretations and virtues of alternative estimation routines and their outputs. This discussion includes the Kalman filter and smoother, and precision based algorithms. As the advantages of using large models have become better understood, a focus has developed on dimension reduction and computational advances to cope with high-dimensional parameter spaces. We give an overview of a number of recent advances in these directions.

Many models suggested by economic theory are either non-linear or non-Gaussian, or both. We discuss work on the particle filtering approach to such models as well as other techniques that use various approximations - to either the time  $t$  state and measurement equations or to the full posterior for the states - to obtain draws.

## **Keywords**

State space model, filter, smoother, non-linear, non-Gaussian, high-dimension, dimension reduction.

## **JEL Classification**

C11, C22, E32

## **Address for correspondence:**

(E) [cama.admin@anu.edu.au](mailto:cama.admin@anu.edu.au)

**ISSN 2206-0332**

[The Centre for Applied Macroeconomic Analysis](#) in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

**The Crawford School of Public Policy** is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

# Bayesian State Space Models in Macroeconometrics

Joshua C.C. Chan                      Rodney W. Strachan<sup>\*†</sup>  
Purdue University and UTS              University of Queensland

October 18, 2020

---

<sup>\*</sup>Corresponding author: [r.strachan@uq.edu.au](mailto:r.strachan@uq.edu.au). Rodney Strachan is a Fellow of the Rimini Centre for Economic Analysis and the Centre for Applied Macroeconomic Analysis.

<sup>†</sup>This research was funded by Australian Research Council Discovery Project grant DP180102373.

## ABSTRACT

State space models play an important role in macroeconometric analysis and the Bayesian approach has been shown to have many advantages. This paper outlines recent developments in state space modelling applied to macroeconomics using Bayesian methods. We outline the directions of recent research, specifically the problems being addressed and the solutions proposed. After presenting a general form for the linear Gaussian model, we discuss the interpretations and virtues of alternative estimation routines and their outputs. This discussion includes the Kalman filter and smoother, and precision based algorithms. As the advantages of using large models have become better understood, a focus has developed on dimension reduction and computational advances to cope with high-dimensional parameter spaces. We give an overview of a number of recent advances in these directions.

Many models suggested by economic theory are either non-linear or non-Gaussian, or both. We discuss work on the particle filtering approach to such models as well as other techniques that use various approximations - to either the time  $t$  state and measurement equations or to the full posterior for the states - to obtain draws.

**Keywords:** State space model, filter, smoother, non-linear, non-Gaussian, high-dimension, dimension reduction.

**JEL Classification:** C11, C22, E32

# 1 Introduction

State space modelling using Bayesian methods has a long history (Kitagawa and Gersch (1984)). These models, also called dynamic linear models, are an incredibly powerful tool for capturing unobserved processes and the specification encompasses a wide range of potential mechanisms. As such, they play an important role in macroeconomic modelling. They have, for example, allowed us to better estimate trends and cycles (Harvey (1985), Watson (1986) and more recently Morley (2002), Harvey, Trimbur and van Dijk (2007), Stock and Watson (2007), Chan, León-González, and Strachan (2018), Grant and Chan (2017)), they arise rather naturally in dynamic stochastic general equilibrium (DSGE) models (e.g., see Kulish, Morley and Robinson (2017) or for an excellent overview of Bayesian estimation of the DSGE see Herbst and Schorfheide (2016)), they have been used to improve forecasting and to account for structural change (Harvey and Todd (1983), Aoki (1987), Clark (2011), D’Agostino, Gambetti, and Giannone (2013)), they help us to understand how uncertainty and volatility have evolved and how the transmission of monetary policy shocks has evolved over time (Cogley and Sargent (2001, 2005), Primiceri (2005), Carriero, Clark and Marcellino (2016, 2018)).

There exists many alternative frameworks for models that evolve, such as non-parametric Bayesian or Gaussian processes and local likelihood approaches (see, for example, Kalli and Griffin (2018) and Petrova (2019)). State space models share features with many of the non-parametric models, but with the notable distinction that they tend to define evolution of the time-varying parameters, or states, in the direction of time (with an exception to this in Koop and Potter (2010)). State space models, however, have a number of advantages such as the ease of developing economically sensible specifications for the evolution of parameters and the existence of efficient

and accessible algorithms for their estimation. For these reasons, a wide range of macroeconomic models have been framed in the state space form. Examples include single equation models such as unobserved components models, multivariate models such as time-varying parameter vector autoregression models (TVP-VAR), and both static and dynamic factor models.

The literature on state space modelling is quite extensive and so, in this article, we focus upon its application in Bayesian macroeconomic analysis. Even with this limitation, the survey will be incomplete as we attempt to focus on only a few of the more recent contributions; although we will necessarily discuss more established ones.

We focus primarily upon the linear Gaussian state space model with some discussion of extensions later in the article. To set the specification and notation, for  $t = 1, \dots, T$ , let  $y_t$  be an  $n \times 1$  vector of variables of interest so that we have a model for a multivariate process, and let  $x_t$  be an  $n \times k$  matrix of regressors. Then a reasonably general form for the linear Gaussian state space model is

$$y_t = x_t \pi_t + \varepsilon_t \tag{1}$$

$$\begin{aligned} \pi_t - \mu &= P(\pi_{t-1} - \mu) + R\eta_t \tag{2} \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim N\left(0, \begin{bmatrix} \Sigma_t & 0 \\ 0 & \Omega_t \end{bmatrix}\right) \text{ and } \pi_1 \sim N(\underline{\pi}_1, V_1). \end{aligned}$$

The dimensions of  $\pi_t$  and  $\mu$  ( $k \times 1$ ) and the known matrix  $P$  ( $k \times k$ ) are determined by the column dimension of the regressor  $x_t$  while the known matrix  $R$  ( $k \times r$ ) need not be square, i.e.,  $r \leq k$ .

The equation (1) is commonly called the *measurement equation* or *observation equation* and provides the link from the data to the states,  $\pi_t$ . The distribution chosen for  $\varepsilon_t$  and (1) for all  $t$  will in most cases define the likelihood. The *state*

*equation* or transition equation (2) governs the evolution of the states or time varying parameters  $\pi_t$ . With the distributions for  $\pi_1$  and  $\eta_t$  in a Bayesian analysis, the state equation defines the prior. Of particular importance is the prior correlation structure this implies for the states.

In macroeconometrics, it is common to assume  $E(\varepsilon_t \eta_t') = 0$  and that  $P$  and  $R$  are  $(k \times k)$  identity matrices;  $P = R = I_k$  and so  $r = k$ , although more general specifications of the state space model have been employed. Cogley and Sargent (2001), for example, extend this specification to allow  $E(\varepsilon_t \eta_t') \neq 0$ . In an unobserved components (UC) model, Morley (2010) uses  $x_t = (1, 1)$ ,  $\Sigma_t = 0 \forall t$ , an AR(2) for the cycle (the first element of  $\pi_t$ ) and permits the shocks to the trend (the second element of  $\pi_t$ ) and cycle to be correlated.

The random walk specification for the state equation ( $P = R = I_k$ )

$$\pi_t = \pi_{t-1} + \eta_t$$

is a popular default. Although it may not always have a sound justification as a prior, this specification does have important virtues such as inducing a greater degree of smoothness than the stationary specification. It may also more closely match our prior beliefs than a stationary specification. In finance, to obtain a realistic model of the variance it is common when specifying the state equation for stochastic volatility (we discuss this in more detail in Section 4) to assume  $P$  is a scalar and  $|P| < 1$ . However, the volatilities for macroeconomic processes are not well represented by stationary, mean reverting processes (Eisenstat and Strachan (2016)). Typical estimates of the volatility of, say, inflation, log real per capita GDP or unemployment, display deviations of long durations. They may be stationary processes but a linear autoregressive process will not adequately capture the behaviour. Eisenstat and Strachan

(2016) make this argument and introduce a change-point stationary model for the state equation which allowed estimation of the time varying mean and persistence of the states.

We structure the remaining sections of this paper as follows. In Section 2, we outline the most popular, almost universally adopted, estimation procedure - the Kalman filter - as well as another procedure that shows promise for high-dimensional systems. In Section 3, we consider approaches to deal with overparameterization. The discussion first covers techniques that allow the data to decide whether each state should vary over time. We then discuss approaches to reducing the dimension of the entire model and thereby permitting estimation of larger models. The applications in both cases are to TVP-VAR models. Section 4 considers popular and recent approaches to estimation of non-linear and non-Gaussian state space models where the Kalman filter cannot be applied. Section 5 concludes with suggestions for future research.

## 2 Estimation of the linear Gaussian model

Estimation of the  $(Tk \times 1)$  vector of states  $\pi = (\pi'_1, \dots, \pi'_T)'$  is most commonly done using the Kalman filter and Kalman smoother. The Kalman filter, first proposed by Kalman (1960) and Kalman and Busy (1961), has been studied extensively and explanations are available in several books and articles; one excellent and very accessible reference is Durbin and Koopman (2001). To begin establishing notation, denote by  $y^{1:t}$  the information available in the data up to time  $t$  such that  $y^{1:T}$  represents all of the information in the sample. Beginning with  $V_1$  and  $\pi_1$ , the Kalman filter produces, at time  $t$ , estimates of the mean and variance of the state  $\pi_t$  using the data up to  $t - 1$ , that is  $p_t = E(\pi_t | y^{1:t-1})$  and  $V_t = V(\pi_t | y^{1:t-1})$ . One representation of the steps

in the Kalman filter is

$$\begin{aligned} \nu_t &= y_t - x_t \pi_t & F_t &= x_t V_t x_t' + \Sigma_t \\ K_t &= V_t x_t' F_t^{-1} & L_t &= I - K_t x_t \\ p_{t+1} &= p_t + K_t \nu_t & V_{t+1} &= V_t L_t' + R \Omega_t R'. \end{aligned}$$

The contemporaneous filtering equations (see Durbin and Koopman (2001) page 68) are another representation of this filtering approach but produce estimates of the mean and variance of the state  $\pi_t$  using the data up to  $t$ , that is  $p_{t|t} = E(\pi_t | y^{1:t})$  and  $V_{t|t} = V(\pi_t | y^{1:t})$ . The Kalman smoother, by contrast, produces estimates of the mean and variance of the state  $\pi_t$  using all available data, that is  $\hat{\pi}_t = E(\pi_t | y^{1:T})$  and  $\hat{V}_t = V(\pi_t | y^{1:T})$  and so from a statistical perspective can be said to be efficient. To obtain the smoothed estimates, computation begins with the end values  $r_T = 0$  and  $N_T = 0$  and works backwards from  $t = T$  to  $t = 1$  through the recursions

$$\begin{aligned} r_{t-1} &= x_t' F_t \nu_t & N_{t-1} &= x_t' F_t^{-1} x_t + L_t' N_t L_t \\ \hat{\pi}_t &= p_t + V_t r_{t-1} & \hat{V}_t &= V_t - V_t N_{t-1} V_t. \end{aligned}$$

There have been substantial improvements in Kalman filter-based algorithms over the past thirty years in estimating linear Gaussian state space models. Important examples include Carter and Kohn (1994), Frühwirth-Schnatter (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002).

There have been many applications of the Kalman filter and Kalman smoother to estimation of models in Bayesian macroeconomics. For example, the sequence of papers of Cogley and Sargent (2001, 2005) and Primiceri (2005) used TVP-VAR to address the important question of changes in the inflation-unemployment relationship and its implications for monetary policy transmission. These papers informed us on

evidence around the evolution of core inflation, the natural rate of unemployment, inflation variance and predictability and the conduct of monetary policy. Baumeister and Benati (2010) use similar models to study the impact on output growth and inflation of a compression in the yield spread induced by central banks' asset purchases when there is a zero lower bound on the policy rate. Baumeister and Peersman (2013) evidence the decline of the short-run price elasticity of oil demand since the mid-1980s. Lubik and Matthes (2015a) compare estimates of the natural rate of interest from a TVP-VAR to the more commonly used estimate obtained from the unobserved components model of Lauback and Williams. A useful introduction to TVP-VAR and its rationale is provided in Lubik and Matthes (2015b). Due to overparameterization concerns, most papers that use quarterly post World War II data consider only small systems with 3-6 variables. Some recently proposed approaches that allow estimation of large VARs using fat data (the number of variables is large relative to the number of observations) will be discussed in Section 3.

When undertaking a study using the state space model, the statistician has the choice to report either filtered estimates -  $p_t$  and  $V_t$  (or  $p_{t|t}$  and  $V_{t|t}$ ) - or smoothed estimates -  $\hat{\pi}_t$  and  $\hat{V}_t$ . Each of these estimators estimates different things in both a statistical and an economic sense, and so each of these estimates will serve different purposes. When using either the filter or smoother in macroeconometrics, it is worthwhile to give some consideration to the economic meaning of the outputs in each case and to what purpose each is best suited. Estimates from the Kalman filter  $p_t$  and  $V_t$  may include variation beyond actual variation in the underlying state and this excess variation is due to learning (Sims (2001)). Such variation is incorporated into the estimates from the Kalman smoother.

In comments on Cogley and Sargent (2001), Sims (2001) pointed out the significance from an economic perspective of the differences in the outputs from the Kalman

filter and the Kalman smoother. The difference between the filtered and smoothed estimates tells us something about the difference between the best estimates made at the time  $t$  and ex-post estimates that use all available data today (or at time  $T$ ). Another less formal way to look at this difference is to think of smoothed estimates as estimating what was actually happening at the time  $t$  whereas filtered estimates replicate estimates using the information available to the econometrician at time  $t$ ; possibly to then replicate forecasts made at that time. The latter view is not entirely satisfactory because it does not take into account that the econometrician's information set at time  $t$  would include unrevised or real-time data, not final data. In this sense it is less clear from an economic perspective how to interpret filtered estimates produced using final data.

The use of the Kalman filter and smoother to estimate the state space model has become so commonplace that the estimation routine and model are sometimes referred to synonymously. It is important to distinguish the model from the estimation, however, as there do exist alternative approaches to estimation. More recently, a new class of algorithms has been proposed for estimating, primarily, linear Gaussian state space models. These are the precision-based algorithms, or precision samplers, and they have been proposed a number of times. For example, Rue (2001), Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011) all propose similar approaches. These estimators do not use a filtering approach, rather they draw all states in one step in much the same way one would draw a vector of regression coefficients from its posterior distribution.

In outlining the precision sampler, we follow the approach of Chan and Jeliazkov (2009). Stacking all states into one large  $(Tk \times 1)$  vector  $\pi = (\pi'_1, \dots, \pi'_T)'$ , the conditional (upon the parameters) posterior for this vector is normal with  $(Tk \times 1)$

mean vector  $\bar{\pi}$  and covariance matrix  $\bar{V}_\pi$ . To see this, stack (1) over time as

$$y = X\pi + \varepsilon \quad \varepsilon \sim N(0, \Sigma)$$

where  $y = (y'_1, \dots, y'_T)'$  and  $\varepsilon = (\varepsilon'_1, \dots, \varepsilon'_T)'$  with dimensions  $(nT \times 1)$  and  $(nT \times 1)$  respectively, and

$$X = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & x_T \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_T \end{bmatrix}.$$

Then stack the state equation (2) over time to obtain

$$D\pi = \underline{\pi} + M\eta \quad \eta \sim N(0, \Omega)$$

$$D = \begin{bmatrix} I_k & 0 & 0 & \cdots & 0 \\ -P & I_k & 0 & & 0 \\ 0 & -P & I_k & \cdots & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & \cdots & -P & I_k \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{T-1} \\ \eta_T \end{bmatrix},$$

$$\Omega = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_T \end{bmatrix}, \quad \underline{\pi} = \begin{bmatrix} \underline{\pi}_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},$$

and  $M = (I_T \otimes R)$ . Assuming  $r = k$  and  $R = I_k$ , Chan and Jeliazkov (2009) show

the resulting posterior mean ( $\bar{\pi}$ ) and covariance matrix are ( $\bar{V}_\pi$ ) defined by

$$\begin{aligned}\bar{V}_\pi &= [D'\Omega^{-1}D + X'\Sigma^{-1}X]^{-1} \\ \bar{\pi} &= \bar{V}_\pi [D'\Omega^{-1}\underline{\pi} + X'\Sigma^{-1}y]\end{aligned}$$

and the full vector of states has posterior  $\pi \sim N(\bar{\pi}, \bar{V}_\pi) = p(\pi|y)$ .

The precision algorithms take advantage of the sparse and banded structure of many of the very large matrices that enter into the posterior for the states. For example, since each of  $D$ ,  $\Omega$ ,  $X$  and  $\Sigma$  are banded - that is, they mostly consist of zeros and the relatively few non-zero elements are arranged along the main diagonal - the matrices  $D'\Omega^{-1}D$  and  $X'\Sigma^{-1}X$  are also banded. As a result, the  $(Tk \times Tk)$  inverse covariance matrix or precision matrix

$$P_\pi = \bar{V}_\pi^{-1} = D'\Omega^{-1}D + X'\Sigma^{-1}X$$

is banded as well. These features, plus the fact that the matrix  $P_\pi$  is symmetric, imply a much lower cost of storage and fewer computations are needed. Importantly, we need not compute the inverse  $\bar{V}_\pi = P_\pi^{-1}$ ; doing so would involve  $\mathfrak{O}(T^3)$  operations. Taking into account this structure and using efficient functions in standard software packages (e.g., Matlab), the algorithms avoid many redundant computations and reduce the order of operations to  $\mathfrak{O}(T)$ . Having  $P_\pi$ , we first take the Cholesky  $P_\pi = C'C$  where  $C$  is an upper-triangular matrix, and write the equations defining the mean vector as

$$C'C\bar{\pi} = D'\Omega^{-1}\underline{\pi} + X'\Sigma^{-1}y.$$

Second, we define  $\gamma = C\bar{\pi}$  and solve by forward substitution the following system of

equations for  $\gamma$ ,

$$C'\gamma = D'\Omega^{-1}\underline{\pi} + X'\Sigma^{-1}y.$$

Finally, having  $\gamma$ , we then solve the system of equations  $C\bar{\pi} = \gamma$  for  $\bar{\pi}$ . Similarly, we can use the same type of operations to obtain a draw of the  $(Tk \times 1)$  vector of states,  $\pi$ . That is, we draw a  $(Tk \times 1)$  vector of independent standard Normal random variables,  $z \sim N(0, I_{Tk})$  and solve  $Cw = z$  for  $w$ . The vector  $\pi = w + \bar{\pi}$  is a draw from  $N(\bar{\pi}, \bar{V}_\pi)$ . These operations result in a lower order of computations to produce each estimated state than the Kalman smoother, but the algorithm produces the same outputs as does the Kalman smoother. The significance of this precision based approach is that, with a lower cost of computation, macroeconometricians are able to consider estimation of larger models. As the precision sampler saves on computations, and this saving increases with dimension  $n$ , they enable estimation of higher dimensional systems for less additional cost. In a recent paper, Chan, Eisenstat and Strachan (2020) take advantage of this in estimating a high dimensional TVP-VAR model. In their specification,  $P = R = I_k$  and  $\Omega_t = \Omega$  has reduced rank. While less common than the Kalman filter and smoother, the precision sampler has begun to be applied such as in the study of evolution of the Phillips curve (Fu (2020)), studying the relative virtues of “commitment” versus “discretion” in the conduct of monetary policy (Du (2016)) and the impact on the Australian business cycle of international shocks (Cross and Poon (2019)).

### 3 Approaches to overparameterization

Macroeconometricians have come to understand the importance, for a range of reasons, of allowing model parameters to evolve over time. One important example of

this is in VAR models with time varying parameters and, in particular, with heteroskedasticity. These models have led to an improved understanding of policy and the macroeconomic environment (see, among many others, Cogley and Sargent (2005), Cogley, Morozov and Sargent (2005), Primiceri (2005), Koop, Léon-González and Strachan (2009) and Canova and Ferroni (2012)). A natural candidate specification to permit evolution of the parameters, and one that is frequently used for time varying parameter models, is the state space model and the TVP-VAR in state space form has proven very successful.

Over recent years, there has been an increasing interest in estimating higher dimensional state space models with the TVP-VAR being an important example. Alongside the literature on time varying parameters has been work pointing out empirical and theoretical issues that arise from using small VARs, that is VARs with only a few variables (Carriero, Kapetanios and Marcellino (2011), Giannone, Lenza, Momferatou and Onorante (2014), Koop (2013) and Koop and Korobilis (2013)). Mapping from a DSGE model to a VAR involves approximations to the implied specification (such as log linearization or truncating the lag length of the VAR( $\infty$ )) (Giacomini (2013)) or encounters the issue of nonfundamentality (resulting in the errors in the DSGE not mapping to the structural errors of the VAR) (Fernández-Villaverde and Rubio-Ramírez (2007)). Moving to a larger VAR will often mitigate these concerns. In the case where we map from a DSGE, using a larger VAR is an effort to better match the information set of the econometrician to that of agents acting in the model. Theoretical issues such as these have naturally led to consideration of how to estimate larger models.

Larger models also have empirical advantages. For example, in forecasting there has been growing evidence that large Bayesian VAR models (BVARs) perform better than factor models. In what is perhaps the seminal paper on using large BVARs,

Banbura, Giannone and Reichlin (2010) use shrinkage techniques to permit estimation and forecasting with 7, 20 and 131 macroeconomic variables. They find, using mean squared forecast error (MSFE), “... that the largest specification outperforms the small models in forecast accuracy and produces credible impulse responses ...”. Usefully, they also find that much of the gain is achieved using the medium sized model with 20 variables. Carriero, Kapetanios and Marcellino (2009) demonstrate improved forecasting performance based upon MSFE and mean absolute forecast error of a large (33 variables) BVAR over the driftless random walk for forecasting exchange rates. Further studies in this area, such as Carriero, Clark and Marcellino (2019), Giannone, Lenza, Momferatou and Onorante (2014), and Koop (2011), continue to accumulate and clarify the evidence on when (e.g., for which variable or for models that include stochastic volatility) and in what sense (e.g., in point forecasts or density forecasts), large BVARs forecast well.

These two streams of literature have naturally led to work developing higher dimensional and time varying models. As the number of variables included in the VAR increases, however, such models quickly become overparameterized or overfitting leading to poor estimation and inference. Allowing for time-variation, with large TVP-VAR models, significantly exacerbates this issue. To mitigate the overparameterization issues, researchers have sought ways to reduce the dimensions of such models to improve estimation.

One approach to achieving parsimony is to allow the data to select parameters to be set to zero: an application of a technique commonly called *variable selection*. A second approach is to allow the distribution of the parameter to be shrunk towards zero, so that the parameter can effectively be treated as if it were zero. This is the *shrinkage* approach. A third approach treats the entire model structure (rather than individual parameters) to reduce the dimension of the model. We leave discussion

of this last approach to the end of this section. These techniques for overcoming overparameterization in principle can be applied to estimate large VARs using fat data (e.g., see Koop (2016)).

Variable selection (see, for example, George and McCulloch (1993) and Smith and Kohn (1996)) involves equating the value of a binary indicator,  $\gamma \in \{0, 1\}$ , to the restriction that a parameter is set to zero or not. This gives a prior distribution for the parameter that is a discrete-continuous mixture: the prior has a spike at zero with a continuous distribution elsewhere and for this reason is commonly called the *spike-and-slab* prior. The aim in using this prior is to empirically determine whether a parameter is exactly zero, or is drawn from a continuous distribution. In the context of a VAR, this variable selection technique can be used to automatically select lag length, to identify the set of variables to be included in each equation or to identify which variables have time varying coefficients and which do not (see, for a recent example, Bitto and Frühwirth-Schnatter (2019)). Shrinkage priors, on the other hand, do not set parameters to zero but place mass around zero and are continuous everywhere. The advantage of shrinkage priors is that they are generally computationally more efficient. The idea in this approach is that parameters that are really zero will have posteriors that are shrunk towards zero. When this occurs, the parameter is treated as *effectively* zero. The researcher must decide between the computational advantages of the shrinkage priors (with no dimension reduction) and the actual dimension reduction (but computational challenges) of using spike-and-slab priors for variable selection.

Gerlach, Carter and Kohn (2000) and Frühwirth-Schnatter and Wagner (2010), use related approaches to let the data decide which parameters are time varying. The more general and earlier approach is that of Gerlach *et al.* (2000) in which the time-variation is allowed to switch on and off where necessary. To do this, they developed

a dynamic mixture approach that could allow time variation of the states at some points in time but to turn off time variation at other times. This approach uses an indicator variable,  $\gamma_t$ , that could turn off the shocks to the states. The state equation is specified as  $\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + \gamma_t \eta_t$  where  $\gamma_t \in \{0, 1\}$ . A value of  $\gamma_t = 0$  implies the state does not vary from time  $t - 1$  to  $t$ . One way to interpret this is that the shock at time  $t$  was zero,  $\eta_t = 0$ . Another interpretation is that the variance of the shock is zero,  $E(\eta_t^2) = 0$ . We will return to this latter interpretation shortly.

As the indicator's value can change over time, computation can prove to be slow or challenging. An important contribution of Gerlach *et al.* (2000) was the development of an efficient algorithm for sampling the indicators by first integrating out the states. The model proposed by Gerlach *et al.* (2000) permits more general specifications than the one we have discussed here and the approach has been applied in a number of macroeconometric studies including Koop *et al.* (2009), Chan, Koop, Léon-González and Strachan (2012), Koop and Korobilis (2012), Groen, Paap and Ravazzolo (2013), Korobilis (2013) and Clark and Ravazzolo (2015).

Frühwirth-Schnatter and Wagner (2010) frame the question of which parameters should vary over time as a variable selection problem. The set up resembles that of Gerlach *et al.* (2000) with the addition of a transformation to the non-centered parameterization, although it is less general in that it does not allow an indicator for each point in time. In Frühwirth-Schnatter and Wagner (2010), the parameter that is to be set to zero if  $\gamma = 0$  is the variance of the state equation in which case the time variation is turned off.

To demonstrate this approach, consider a scalar state and a standard form of the measurement and state equations as discussed above with respect to the approach of

Gerlach *et al.* (2000), but with a single time-invariant indicator

$$\begin{aligned} y_t &= \tilde{\alpha}_t + \varepsilon_t \\ \tilde{\alpha}_t &= \tilde{\alpha}_{t-1} + \gamma\eta_t \end{aligned}$$

where  $\gamma \in \{0, 1\}$ ,  $\eta_t \sim N(0, \tilde{\sigma}^2)$  (and so we could roughly write  $\gamma\eta_t \sim N(0, \gamma^2\tilde{\sigma}^2)$ ) and  $\tilde{\alpha}_1 = \alpha \sim N(0, \omega^2)$ . By subtracting the initial state from all states and dividing this by  $\gamma\tilde{\sigma}$  (we ignore the obvious problem of dividing by zero when  $\gamma = 0$  to preserve the conceptual link between Gerlach *et al.* (2000) and Frühwirth-Schnatter and Wagner (2010)) we have the non-centered parameterization,  $\tilde{\alpha}_t = \alpha + \gamma\tilde{\sigma}\alpha_t$  and the model becomes

$$\begin{aligned} y_t &= \alpha + \gamma\tilde{\sigma}\alpha_t + \varepsilon_t \\ \alpha_t &= \alpha_{t-1} + z_t \end{aligned}$$

where  $z_t \sim N(0, 1)$  and  $\alpha_1 = 0$ . The constraint  $\tilde{\sigma} > 0$  is treated as an identification issue in Frühwirth-Schnatter and Wagner (2010) although here we provide a different interpretation and parameterize this identification issue. To see this, introduce an unidentified bivariate random variable  $\iota \in \{-1, 1\}$  where  $P(\iota = 1) = 1 - P(\iota = -1)$  and insert this into the measurement equation as  $\tilde{\sigma}\alpha_t = \tilde{\sigma}\iota\alpha_t = \sigma\alpha_t$  where  $\sigma = \iota\tilde{\sigma}$ . The state equation is now  $y_t = \tilde{\alpha}_0 + \gamma\sigma\alpha_t + \varepsilon_t$ . Although  $\alpha_t$  has been transformed we do not change the notation as it has the same distribution before and after the transformation. This transformation using the unidentified  $\iota$  is a form of parameter expansion (see discussions in Liu, Rubin and Wu (1998) and Liu and Wu (1999)) that is commonly used to improve computation. An important requirement for a parameter expansion to improve computation is that the prior distribution of the

unidentified parameter ( $\iota$ ) is independent of the parameters in the model, which it is in this case.

The state equation error standard deviation,  $\sigma$ , is now just a regression coefficient on the state with support  $\mathbb{R}$ . If the state is not varying over time then we would expect to see this coefficient restricted to  $\sigma = 0$ , restricting the ‘variable’  $\alpha_t$  out of the equation. To implement variable selection the binary indicator  $\gamma \in \{0, 1\}$  in the measurement equation is given a prior,  $P(\gamma = 1) = P(\gamma = 0) = \frac{1}{2}$  say. Thus we see the link to Gerlach *et al.* (2000) as  $y_t = \tilde{\alpha}_0 + \gamma\sigma\alpha_t + \varepsilon_t$ . If  $\gamma = 0$  then  $\sigma = 0$  and the state is constant, whereas if  $\gamma = 1$  then  $\sigma \neq 0$ , and the state may vary over time. It is for this reason Bitto and Frühwirth-Schnatter (2019) refer to this as a variance selection problem. The posterior probability  $P(\gamma = 0|y)$  tells us the empirical support for the restriction that the state is time invariant.

There have been a number of applications and extensions of the non-centered parameterization for state space models in macroeconometrics. Chan (2018), for example, uses this specification with the Savage-Dickey density ratio to compute the Bayes factor for the restriction to time-invariance of parameters in various models of inflation. Towards the end of this section we discuss an extension and generalization of the specification in Frühwirth-Schnatter and Wagner (2010) by Chan, Eisenstat and Strachan (2020).

Despite its strong theoretical appeal, variable selection tends to be more computationally difficult than the approximation to variable selection through shrinkage. For this reason there have been many approaches inducing shrinkage including Belmonte, Koop, and Korobilis (2014), Kalli and Griffin (2014), Uribe and Lopes (2017), Rockova and McAlinn (2017), Huber, Koop and Onorante (2020) and Bitto and Frühwirth-Schnatter (2019). In the following discussion we consider a few implementations of shrinkage.

An important contribution of Frühwirth-Schnatter and Wagner (2010) is the argument for using a normal prior on the parameter  $\sigma$  rather than the traditional default of the inverted gamma on  $\tilde{\sigma}^2$ . The difference, which is important for questions of parsimony, is in how these two priors treat values near zero. The more commonly used inverted gamma distribution on  $\tilde{\sigma}^2$  prevents shrinkage as it is bounded away from zero. The normal prior on  $\sigma$ , which implies a gamma prior on  $\sigma^2$ , is more innocuous in that it allows weight around zero. In an inflation forecasting exercise, Belmonte *et al.* (2014) remove the variable selection parameter ( $\gamma$ ) and use shrinkage priors on the prior variance of the initial state ( $\tilde{\alpha}_0$ ) and the state variance ( $\tilde{\sigma}$ ) in the non-centered specification. The prior uses a Lasso prior for the standard deviation.<sup>1</sup>

Bitto and Frühwirth-Schnatter (2019) take a similar approach to Belmonte *et al.* (2014) but use a Normal-gamma prior for shrinkage which encompasses the Lasso prior as a special case. The prior in Bitto and Frühwirth-Schnatter (2019) can produce a more pronounced peak at zero while allowing considerable mass away from zero, effectively bringing the continuous shrinkage approach closer to the variable selection or spike-and-slab approach. Giannone, Lenza and Primiceri (2018) explore the support for sparse (as could result from spike-and-slab priors) versus dense (models with many parameters shrunk towards zero) time invariant models. They find support for dense models over sparse models. A paper that brings together selection and shrinkage is Huber *et al.* (2020) using *sparsification* (see Barbieri and Berger (2004)) where sparsification involves setting parameters with low posterior probabilities of being non-zero to zero. In a TVP-VAR, and in contrast to the results of Giannone *et al.* (2018), they find strong evidence of sparsity and that large sparse models forecast well.

---

<sup>1</sup>The Lasso of Tibshirani (1996) can be formulated as a penalized regression with  $\ell^1$  penalty. As pointed out in Tibshirani (1996), this penalty can be interpreted as a Laplace prior distribution. For more details, see, e.g., Chan *et al.* (2019), pp.333-337.

Another approach to computing high dimensional state space models is to change the model specification in such a way that results in significant computational and memory savings. We discuss three such approaches. The first of these makes use of techniques developed over fifty years ago when computing power was very limited. In a series of papers Koop and Korobilis (2011, 2012, 2013) (among others) apply this technique in a *dynamic model averaging* approach suggested by Raftery, Karny and Ettlter (2010). The updating equation for the filtered estimate of the state covariance matrix,  $V_{t+1}$ , in the contemporaneous filtering equations representation is

$$V_{t+1} = V_{t|t} + R\Omega_t R'.$$

Raftery *et al.* (2010) introduce the *forgetting factor*  $\lambda$ , where  $0 < \lambda \leq 1$ , and replace the above updating equation with

$$V_{t+1} = \frac{1}{\lambda} V_{t|t}.$$

If we assume  $\Omega_t = (\lambda^{-1} - 1) V_{t|t}$ , then this is the correct updating equation for a properly defined state space model. Using forgetting factors to alter the model specification in this way significantly reduces the computational burden. This computational advantage comes from not needing to compute or simulate  $\Omega_t$ . Koop and Korobilis (2013) clearly demonstrate this advantage when they compute a TVP-VAR with four lags for 25 variables, implying a state vector with dimension  $k = 2,525$ . They demonstrate that this is particularly useful in a forecasting exercise.

Koop and Korobilis (2013) avoid a full exploration of the posterior distribution which is appropriate for a forecasting exercise. However, there are important questions that do require knowledge of the full posterior. For example, simple measures of un-

certainty are the highest posterior density or the percentile range (e.g., inter-quartile range) for objects of interest. Chan, Eisenstat and Strachan (2020) present an approach that provides a full MCMC exploration of the posterior for a high-dimensional TVP-VAR with stochastic volatility (TVP-VAR-SV). This approach also imposes a change in the model specification but one that is suggested by the data. That is, they impose rank reduction on the state equation covariance matrix  $\Omega$  such that they estimate a singular state space model. This feature of the state covariance matrix was noted in Cogley and Sargent (2005). This simple and empirically supported restriction results in a reduction of the dimension of the model by around 90% to 99%, thereby making the model computable. Chan *et al.* (2020) improve computation by implementing a number of techniques: the rank restriction, they estimate the VAR in structural rather than reduced form, generalize the non-centered parameterization of Frühwirth-Schnatter and Wagner (2010) to the matrix case, employ parameter expansions and use the precision sampling approach of Chan and Jeliazkov (2009). They apply this to a study of the evolution of the transmission of news and non-news shocks in a system with 15 variables.

The approach in Chan *et al.* (2020) works by reducing the dimension of the parameter space by reducing the sources of variation in the states. Koop, Korobilis and Pettenuzzo (2019) achieve a reduction in the dimension of the parameter space by using *compressed regression* to reduce the dimension of the data. Bayesian compressed regressions, introduced by Guhaniyogi and Dunson (2015), randomly project the data into a lower dimension. Importantly, the draws of these projections are not informed by the data. To demonstrate this in the linear regression model for a  $(T \times 1)$  vector  $y$  with a  $T \times k$  matrix of regressors  $X$ , the compression is applied as follows

$$y = X\beta + \varepsilon = X\Phi\gamma + \varepsilon = X^\Phi\gamma + \varepsilon$$

where  $X^\Phi = X\Phi$ ,  $\gamma$  is an  $(q \times 1)$  vector and  $\Phi$  is the  $(k \times q)$  ( $q < k$ ) random projection matrix. Each element of  $\Phi$  depends upon a single scalar parameter,  $\nu$ , which is drawn independently of the data. The  $(i, j)^{th}$  element of  $\Phi$ ,  $\Phi_{ij}$ , takes one of three discrete values determined by the probability of each value. Specifically,  $\Phi_{ij} \in \left\{ -\frac{1}{\sqrt{\nu}}, 0, \frac{1}{\sqrt{\nu}} \right\}$  with probabilities  $\nu^2$ ,  $2(1 - \nu)\nu$  and  $(1 - \nu)^2$  respectively. The unknown parameter  $\nu$  might be drawn, for example, from a uniform distribution bounded away from zero and one for numerical stability. When choosing  $q$ , Guhaniyogi and Dunson (2015) suggest using Bayesian Model Averaging. Koop *et al.* (2019) apply this to the TVP-VAR by combining it with dynamic model averaging. By using the structural form of the VAR, they apply compression to each equation one at a time and the computational result is impressive. They are able to apply this to a system with  $n = 129$  variables and 13 lags. The state vector for a full TVP-VAR with these specifications would have dimension over  $k > 210,000$ . The approach is demonstrated to be empirically useful in a forecasting exercise in which it performs well. Zhao, Zie and West (2016) and McAlinn and West (2019) outline the application of approaches similar to that of Koop and Korobilis (2013) for forecasting in finance and macroeconomics. The evidence in Koop *et al.* (2019) strongly suggests that large, time-varying parameter models perform well in point forecasts and supports the view developing in the literature that the addition of stochastic volatility significantly improves density forecasts.

In this section we have discussed various approaches to dealing with parameter proliferation which often occurs in state space models used in macroeconomics. The range of approaches involve shrinkage, variable selection or alterations to the structure of the model to either reduce its dimension or permit easier computation. As models grow and the advantages of large models become more apparent and better understood, this is likely to remain a very active area of research.

## 4 Non-linear and non-Gaussian models

An issue that arises in a range of models is that the states  $\pi_t$  evolve nonlinearly, enter the measurement equation non-linearly and/or  $\varepsilon_t$  or  $\eta_t$  are non-Gaussian. To allow for these possibilities, we write the measurement and state equations in a more general form as

$$y_t = F(\pi_t, \varepsilon_t) \text{ for } t > 1 \quad (3)$$

$$\pi_t = H(\pi_{t-1}, \eta_t) \text{ for } t > 1 \quad (4)$$

$$\pi_1 \sim f(\pi_1) \quad (5)$$

or, slightly more generally again,

$$y_t \sim p(y_t | \pi_t) \text{ for } t > 1 \quad (6)$$

$$\pi_t \sim f(\pi_t | \pi_{t-1}) \text{ for } t > 1 \quad (7)$$

$$\pi_1 \sim f(\pi_1). \quad (8)$$

In these cases the Kalman filter and smoother cannot be employed or require extensions to enable inference. A number of alternative approaches have been proposed and in this section we discuss three such approaches. The first involves respecifying the model into a linear Gaussian form such that the Kalman filter can be used. Next we outline an alternative filtering algorithm that allows for very general forms for the state space model, the particle filter. Lastly, we discuss another approach that draws upon the precision sampler approach discussed in Section 2.

In some cases, it is possible to respecify a non-linear and non-Gaussian model into a linear Gaussian form to permit application of standard estimation procedures.

An example is the stochastic volatility model. A simple specification (e.g., Jacquier, Polson and Rossi (1994) or Kim, Shephard and Chib (1998)) of the measurement and state equations for stochastic volatility is

$$\begin{aligned} y_t &= e^{h_t/2} \varepsilon_t, \\ h_t &= \mu + \rho h_{t-1} + \eta_t, \end{aligned}$$

where  $\varepsilon_t \sim iidN(0, 1)$ . The measurement equation is non-linear in the state,  $h_t$ , but Gaussian. Kim *et al.* (1998) apply a sequence of transformations to obtain a linear and (conditionally) Gaussian form. They begin by squaring both sides and then taking the log. The actual transformation of  $y_t$  is  $y_t^* = \ln(y_t^2 + c)$  where  $c > 0$  is a small number added to induce stability. The resulting measurement equation is now

$$y_t^* = h_t + \varepsilon_t^* \text{ where } \varepsilon_t^* = \ln(\varepsilon_t^2) \sim \ln(\chi_1^2).$$

The error  $\varepsilon_t^*$  is distributed as  $\ln(\chi_1^2)$  which is clearly non-Gaussian. The next step is to approximate the distribution  $p(\varepsilon_t^*)$  by a mixture of normals,  $p(\varepsilon_t^*) \cong \sum_{i=1}^7 w_i N(\mu_i, \sigma_i^2)$ . By drawing from each of the seven normals with probability  $w_i$  the overall draws are approximately draws from  $p(\varepsilon_t^*)$ . To implement this, a random latent state is introduced for each point in time  $t$ ,  $s_t = i$  for  $i = 1, \dots, 7$  and  $\Pr(s_t = i) = w_i$ . Conditional upon the drawn state  $s_t = i$ ,  $\varepsilon_t^* \sim N(\mu_i, \sigma_i^2)$  and the model is now linear and Gaussian which permits the use of the Kalman filter or smoother. Introducing the latent state or auxiliary variable  $s_t$  results in a significant simplification of the sampling algorithm. This approach - the auxiliary mixture sampling - has been applied widely. Examples include Frühwirth-Schnatter and Wagner (2006), Omori, Chib, Shephard and Nakajima (2007), Frühwirth-Schnatter, Frühwirth and Rue (2009), Barra, Borowska and

Koopman (2018) and Mao and Zhang (2018).

In cases in which the measurement or state equations are non-linear, and such a transformation to a linear form is not possible or practical, then the Extended Kalman Filter (EKF) (Anderson and Moore (1979), Jazwinski (1970)) has often proven useful. The EKF essentially involves taking locally linear approximations to the non-linear functions  $F$  and  $H$ . This linearization uses first order Taylor series expansions of  $F$  and  $H$  to develop linear approximations to these functions which are then used in Gaussian approximations to the predictive distributions,  $p(\pi_t|y^{1:t-1})$  and  $p(\pi_t|y^{1:t})$ . As with any approach, a number of issues have been identified with the EKF (Creal (2012)). For example, if the true predictive distributions are not well approximated by Gaussian distributions (e.g., if they are bimodal or heavily skewed) then the EKF will not perform well (Arulampalam, Maskell, Gordon, and Clapp (2002)). In the case of non-linear measurement and state equations with possibly non-Gaussian shocks, the particle filter approach can be used and has some advantages over the EKF in that there is no need for functional approximation.

Particle filters or sequential Monte Carlo samplers (Del Moral, Doucet and Jasra (2006) and Doucet, de Freitas and Gordon (2001)) are a powerful tool for estimating a very wide range of processes and so have proven useful in many recent applications. The price to pay for this power and flexibility is computational burden, but this cost has been decreasing or, equivalently the computational power has been increasing. This latter phenomenon along with advances in algorithms, have made particle filters increasingly viable. Accessible tutorials on particle filters are given in Arulampalam *et al.* (2002) or, more recently, Doucet and Johansen (2011) and a recent survey by Creal (2012).

The particle filter uses importance sampling and resampling to draw from an approximation to a sequence of probability distributions. Unlike the EKF which uses an

approximation of  $F$  and  $H$ , the particle filter approximates the distributions  $p$  and  $f$ . This has advantages in cases where the local linearization and Gaussian approximation poorly match the true distribution. Another useful feature of this technique is that the probability distributions need only be known up to the normalizing constant and are approximated by a ‘cloud’ of weighted random samples. The weights  $w_t$  are attached to each sampled value to approximate the density of the states. Sequential Monte Carlo methods are employed to propagate these samples, or particles, over time. An empirically and theoretically important problem with sequential importance sampling is that the mass is degenerate. That is, as the algorithm progresses, one weight converges to one while the other weights collapse to zero. A solution to this issue using a Bootstrap filter was proposed by Gordon, Salmon and Smith (1993). These authors propose a resampling step to obtain equally weighted particles. With independent propagation the degeneracy problem is avoided.

To give a feel for the flexibility of the particle filtering approach, recall the state space model in a very general form (6)-(8), one that is potentially non-Gaussian and non-linear. Applying the particle filter at step  $t$  involves drawing a number of potential states  $\pi_t$ , keeping those that are most likely or probable, and then propagating forward to  $\pi_{t+1}$ . A simple version of the Bootstrap filter of Gordon *et al.* (1993) begins with  $m$  draws of the initial state  $\pi_1$ .

1. *Initialize:*

- For  $i = 1, \dots, m$  draw  $\pi_1^{(i)} \sim f(\pi_1)$  and set  $w_1^{(i)} = \frac{1}{m}$ .

For  $t = 2, \dots, T$

2. *Importance sampling:*

- For  $i = 1, \dots, m$  draw  $\pi_t^{(i)} \sim f(\pi_t | \pi_{t-1}^{(i)})$  and set  $\tilde{\pi}_{1:t}^{(i)} = (\pi_{1:t-1}^{(i)}, \pi_t^{(i)})$

- For  $i = 1, \dots, m$  compute and normalize the weights  $w_t^{(i)} \propto p(y_t | \pi_t^{(i)})$ .

### 3. Resampling:

- Resample  $m$  particles  $\{\pi_t^{(i)}\}_{i=1}^m$  with replacement from  $\{\pi_t^{(i)}, w_t^{(i)}\}_{i=1}^m$
- Return to Step 2.

From the notation used to demonstrate the particle filter, it is clear that there is no theoretical constraint on the dimension of the state  $\pi_t$ . In practice, however, the approximation error grows exponentially in the dimension of the state (Rebeschini and van Handel (2015)). There is generally much greater variation in the weights with higher dimension state vectors and so one particle will take almost all of the weight; leading again to a degeneracy issue. There have been a number of papers developing methods to address degeneracy in high dimensional filtering methods (see, for example, Djurić and Bugallo (2013), Naesseth, Lindsten and Schön (2015), Rebeschini and van Handel (2015), Beskos, Crisan, Jasra, Kamatani and Zhou (2017), Robert and Künsch (2017) and Wigren, Murray and Lindsten (2018)). However, there remain few applications of such techniques to what could be regarded as high-dimensional states in macroeconomics.

Above we mentioned the advantage of approximating the distributions rather than the functions in the model and in Section 2 we discussed the computational advantages of the precision sampler (in which we draw all states at once) over the filtering approach. This computational gain becomes more important if we take into account that there is clearly a need for modelling high-dimensional systems in macroeconomics. Chan and Strachan (2012) develop an approach to drawing states from a possibly non-linear and non-Gaussian state space model of (moderately) high dimension. Their application is to evaluation of credit shocks near the zero lower bound

for interest rates. For this application, they estimate a TVP-VAR-SV where the non-linearity and non-Gaussianity are induced by the truncated support of interest rates at zero. Using second order expansions and Gaussian approximations to the likelihood  $p(y|\pi)$  and the density for  $p(\pi)$  implied by the state equations, they develop an approximation to the posterior for the states,  $p(\pi|y)$ . They then develop three sampling schemes for drawing the states from the posterior: a Metropolis-Hastings algorithm, an Accept-Reject Metropolis-Hastings algorithm and finally a collapsed sampler using the cross-entropy method (e.g., Chan and Kroese (2012)) to develop the proposal density.

This approach of estimating non-linear state space models has been used in various forms in a number of applications. Chan, Koop and Potter (2013 and 2016) introduce bounds on trend inflation and unemployment to capture more economically sensible representations that result from, for example, central bank targets for inflation or that unemployment cannot be negative. Lanne and Luoto (2017) apply this approach to obtain improved estimation of the new Keynesian Phillips curve and found strong evidence for noncausality. More recent applications include Cross and Poon (2019), Chan (2020), Dimitrakopoulos and Kolossiatis (2020) and Hou (2020).

## 5 Summary and conclusion

In this paper, we have provided a review of recent developments in state space modelling in macroeconomics taking a Bayesian approach. For the purpose of estimating a linear Gaussian state space model, the Kalman filter remains the most popular choice for analysts. The precision based samplers show promise as an alternative to the Kalman smoother, particularly when economists prefer larger models for empirical or theoretical reasons, however, a return of dynamic model averaging using

forgetting factors shows considerable potential for estimation of large models. These latter models have been shown to perform very well in forecasting exercises. While we found only application of the Kalman filter using forgetting factors, there does not appear to be any impediment to applying the smoother with this approach and doing so could resolve some interpretational issues as well as broaden its application. As the rationale for large models has strengthened, the focus on computational costs has increased leading to approaches that save on computation by dimension reduction - such as turning off time-variation - or better algorithms, or both. Advances in the theory around particle filters has improved inference in non-linear and non-Gaussian models but challenges remain when applying this approach in high-dimensions. A precision based approximation to the posterior for the states, however, shows promise in some applications.

Recent interest has turned to large models, including time-varying parameter models where state space models dominate. There have been considerable advances into computation, specification and inference in these high-dimensional models and this research continues. Challenges, and so new directions for state space modelling, include estimation and inference in time varying parameter vector autoregressive-moving average models (TVP-VARMA) and treatment of identification in large factor models; both of which have theoretical justifications when the underlying model is assumed to be a DSGE model. This takes us beyond using these models only for forecasting and evidence on time variation. Once VARs and vector error correction models (VECM) were established as workhorse models, research turned to consider restrictions on these models as a way to establish support for economic theories or features implied by economic theories. Similarly, evidence on time-variation in the support for restrictions (see, for example, Koop, León-González and Strachan (2010, 2011)) offers further opportunities for new applied work.

## 6 References

- Anderson, B. D. and Moore, J. B. 1979. Optimal Filtering. Prentice-Hall, New Jersey.
- Aoki, M. 1987. State Space Modeling of Time Series. New York: Springer Verlag.
- Arulampalam, S., Maskell, S., Gordon, N. and Clapp, T. 2002. A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174-188.
- Banbura, M., Giannone, D. and Reichlin, L. 2010. Large Bayesian VARs. *Journal of Applied Econometrics* 25: 71-92.
- Barbieri, M. and Berger, J. 2004. Optimal predictive model selection. *Annals of Statistics* 32: 870-897.
- Barra, I., Borowska, A. and Koopman, S.J. 2018. Bayesian dynamic modeling of high-frequency integer price changes. *Journal of Financial Econometrics* 16(3): 384-424.
- Baumeister, C. and Benati, L. 2010. Unconventional Monetary Policy and the Great Recession - Estimating the Impact of a Compression in the Yield Spread at the Zero Lower Bound (October 11, 2010). ECB Working Paper No. 1258.
- Baumeister, C., and Peersman, G. 2013. Time-Varying Effects of Oil Supply Shocks on the US Economy. *American Economic Journal: Macroeconomics*, 5 (4): 1-28.
- Belmonte, M.A.G., Koop, G. and Korobolis, D. 2014. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting* 33: 80–94.
- Beskos, A., Crisan, D., Jasra, A., Kamatani K. and Zhou Y. 2017. A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied*

*Probability* 49: 24-48.

Bitto, A., Frühwirth-Schnatter, S. 2019. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210(1): 75-97.

Canova, F. and Ferroni, F. 2012. The dynamics of US inflation: Can monetary policy explain the changes? *Journal of Econometrics* 167(1): 47-60.

<https://EconPapers.repec.org/RePEc:eee:econom:v:167:y:2012:i:1:p:47-60>.

Carriero A., Clark, T.E., and Marcellino, M. 2016. Common Drifting Volatility in Large Bayesian VARs. *Journal of Business & Economic Statistics*, 34(3): 375-390.

Carriero A., Clark, T.E., and Marcellino, M. 2018. Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics* 100(5): 799-815.

Carriero, A., Clark, T. E. and Marcellino, M. 2019. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics* 212(1): 137-154.

Carriero, A., Kapetanios, G. and Marcellino, M. 2009. Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting* 25(2): 400-417.

Carriero, A., Kapetanios, G., and Marcellino, M. 2011. Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5): 735-761.

Carter, C., and Kohn, R. 1994. On Gibbs sampling for state space models, *Biometrika*, 81: 541-553.

Chan, J.C.C. 2018. Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews* 37(8): 807-823.

Chan, J.C.C. 2020. Large Bayesian VARs: a flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics*, 38(1): 68-79.

- Chan, J.C.C., Eisenstat, E. and Strachan, R. W. 2020. Reducing the state space dimension in a large TVP-VAR. *Journal of Econometrics*, forthcoming.
- Chan, J.C.C. and Jeliazkov, I. 2009. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical modelling and Numerical Optimisation* 1: 101–120.
- Chan, J.C.C., Koop, G., Poirier, D.J. and Tobias, J.L. 2019. Bayesian Econometric Methods (Second Edition). Cambridge University Press.
- Chan, J.C.C., Koop, G. and Potter, S. 2013. A new model of trend inflation. *Journal of Business and Economic Statistics* 31(1): 94-106.
- Chan, J.C.C., Koop, G. and Potter, S. 2016. A bounded model of time variation in trend inflation, NAIRU and the Phillips curve. *Journal of Applied Econometrics* 31(3): 551-565.
- Chan, J.C.C., Koop, G., León-González, R., and Strachan, R.W. 2012. Time varying dimension models. *Journal of Business and Economic Statistics*, 30(3): 358-367
- Chan, J.C.C., Kroese, D.P. 2012. Improved cross-entropy method for estimation. *Statistics and Computing* 22, 1031–1040. <https://doi.org/10.1007/s11222-011-9275-7>
- Chan, J.C.C., R. León-González, R., and Strachan, R.W. 2018. Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association* 113(522): 819-828, DOI: 10.1080/01621459.2017.1287080.
- Chan J.C.C. and Strachan, R.W. 2012. Efficient estimation in non-linear non-Gaussian state-space models. Working paper, Research School of Economics, Australian National University.
- Clark, T.E. 2011. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics* 29(3):327–341.

- Clark, T.E. and Ravazzolo, F. 2015. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics* 30: 551-575.
- Cogley, T., Morozov, S. and Sargent, T. 2005. Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system, *Journal of Economic Dynamics and Control* 29(11): 1893-1925.  
<https://EconPapers.repec.org/RePEc:eee:dyncon:v:29:y:2005:i:11:p:1893-1925>.
- Cogley, T. and Sargent, T.J. 2001, Evolving post World War II U.S. inflation dynamics, *NBER Macroeconomics Annual* 16: 331-373.
- Cogley, T. and Sargent, T.J. 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.. *Review of Economic Dynamics* 8: 262-302.
- Creal, D. 2012. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews* 31(3): 245-296, DOI: 10.1080/07474938.2011.607333
- Cross, J.L. and Poon, A. 2019. On the contribution of international shocks in Australian business cycle fluctuations. *Empirical Economics*, forthcoming, DOI: 10.1007/s00181-019-01752-y
- D'Agostino, A., Gambetti, L. and Giannone, D. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28(1), 82-101.
- de Jong, P., & Shephard, N. 1995. The simulation smoother for time series models. *Biometrika*, 82(2): 339-350, doi:10.2307/2337412.
- Del Moral, P., Doucet, A., and Jasra, A. 2006. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B*, 68: 411-436.
- Dimitrakopoulos, S. and Kolossiatis, M. 2020. Bayesian analysis of moving average stochastic volatility models: modeling in-mean effects and leverage for financial time

series. *Econometric Reviews* 39(4): 319-343.

Djurić, P.M. and Bugallo, M.F. 2013. Particle filtering for high-dimensional systems. *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 352–355.

Doucet, A. and Johansen, A.M. 2011. A tutorial on particle filtering and smoothing: Fifteen years later. in *Handbook of Nonlinear Filtering* (eds. D. Crisan et B. Rozovsky), Oxford University Press.

Doucet, A., de Freitas, N. and Gordon, N.J. (eds.) 2001. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Du, Q. (2016). Monetary Policy Based on Stochastic Model Predictive Control. *Canadian Social Science*, 12 (4), 1-7.

<http://dx.doi.org/10.3968/8318>.

Durbin, J. and Koopman, S. 2001, *Time Series Analysis by State Space Methods* (Oxford University Press, Oxford).

Durbin, J. and Koopman, S. 2002, A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, 89, 603-616.

Eisenstat, E. and Strachan, R.W. 2016. Modelling inflation volatility *Journal of Applied Econometrics* 31(5): 805–820.

Fernández-Villaverde, J., and Rubio-Ramírez, J. 2007. Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies* 74: 1059-1087.

Frühwirth-Schnatter, S. 1994. Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15:183–202.

Frühwirth-Schnatter, S., Frühwirth, H. and Rue, H. 2009. Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*

19(4): 479.

Frühwirth-Schnatter, S. and Wagner, H. 2006. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93(4): 827-841.

Frühwirth-Schnatter, S. and Wagner, H. 2010. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics* 154(1): 85-100.

Fu, B. 2020. Is the slope of the Phillips curve time-varying? Evidence from unobserved components models. *Economic Modelling* 88: 320–340.

George, E. I. and McCulloch, R.E. 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423): 881-889.

Gerlach, R., Carter, C., and Kohn, R. 2000. Efficient Bayesian inference in dynamic mixture models. *Journal of the American Statistical Association* 95: 819–828.

Giacomini, R. 2013. The Relationship Between DSGE and VAR Models. *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics, Vol. 32)*, Emerald Group Publishing Limited: 1-25. [https://doi.org/10.1108/S0731-9053\(2013\)0000031001](https://doi.org/10.1108/S0731-9053(2013)0000031001)

Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. 2014. Short-term inflation projections: A Bayesian vector autoregressive approach. *International Journal of Forecasting* 30 (3): 635-644. <https://EconPapers.repec.org/RePEc:eee:intfor:v:30:y:2014:i:3:p:635-644>.

Giannone, D., Lenza, M. and Primiceri, G. E. 2018. Economic Predictions with Big Data: The Illusion of Sparsity. *FRB of New York Staff Report* No. 847. <http://dx.doi.org/10.2139/ssrn.3166281>

- Gordon N.J., Salmond D.J. and Smith A.F.M. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE-Proceedings-F* 140: 107-113.
- Grant A. and Chan, J.C.C. 2017. Reconciling output gaps: unobserved components model and Hodrick-Prescott filter. *Journal of Economic Dynamics and Control* 75: 114-121.
- Groen, J., Paap, R., Ravazzolo, F. 2013. Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics* 31: 29–44.
- Guhaniyogi, R., Dunson, D. 2015. Bayesian compressed regression. *Journal of the American Statistical Association*. 110: 1500–1514.
- Harvey, A.C. 1985. Trends and Cycles in Macroeconomic Time Series. *Journal of Business and Economic Statistics* 3: 216-227.
- Harvey, A.C. and Todd, P.H.J. 1983. Forecasting economic time series with structural and Box-Jenkins models: A case study. *Journal of Business and Economic Statistics*, 1: 299-307.
- Harvey, A.C., Trimbur T. and van Dijk H.K. 2007. Cyclical components in economic time series: a Bayesian approach. *Journal of Econometrics* 140: 618-49.
- Herbst E. and Schorfheide, F. 2016. Bayesian Estimation of DSGE Models. *Princeton University Press*.
- Hou, C. 2020. Time-varying relationship between inflation and inflation uncertainty. *Oxford Bulletin of Economics and Statistics* 82(1): 83-124.
- Huber, F., Koop, G. & Onorante, L. 2020. Inducing Sparsity and Shrinkage in Time-Varying Parameter Models. *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2020.1713796

- Jacquier, E., Polson, N.G., and Rossi, P.E. 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12: 371–417.
- Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York.
- Kalli M. and Griffin J. 2014. Time-varying sparsity in dynamic regression models. *Journal of Econometrics* 178:2, 779-793.
- Kalli M. and Griffin J. 2018. Bayesian nonparametric vector autoregressive models. *Journal of Econometrics* 203(2): 267-282.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82: 35–45. doi:10.1115/1.3662552.
- Kalman, R.E.; Bucy, R.S. 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering* 83: 95–108. doi:10.1115/1.3658902.
- Kim, S., Shephard, N., and Chib, S. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65: 361-93.
- Kitagawa, G. and Gersch, W. 1984. A Smoothness Priors-State Space Modeling of Time Series with Trend and Seasonality. *Journal of the American Statistical Association* 79: 378-389.
- Koop, G. 2013. Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics* 28: 177-203.
- Koop, G. 2016. Bayesian methods for fat data. Manuscript available at: <https://sites.google.com/site/garykoop/research>.
- Koop, G. and Korobilis, D. 2011. UK Macroeconomic Forecasting with Many Predictors: Which Models Forecast Best and When Do They Do So? *Economic Modelling* 28, 2307-2318.

- Koop, G. and Korobilis, D. 2012. Forecasting Inflation using Dynamic Model Averaging. *International Economic Review*, 53, 867-886.
- Koop, G. and Korobilis, D. 2013. Large Time-varying Parameter VARs. *Journal of Econometrics* 177: 185-198.
- Koop, G., Korobilis, D. and Pettenuzzo, D. 2019. Bayesian Compressed Vector Autoregressions. *Journal of Econometrics* 210: 135-154
- Koop, G., León-González, R. and Strachan, R. W. 2009. On the Evolution of Monetary Policy. *Journal of Economic Dynamics and Control* 33, 997-1017.
- Koop, G., León-González, R. and Strachan R. W. 2010. Dynamic Probabilities of Restrictions in State Space Models: An Application to the Phillips Curve. *Journal of Business and Economic Statistics* 28, 370–379.
- Koop, G., León-González, R. and Strachan R.W. 2011. Bayesian Inference in the Time Varying Cointegration Model. *The Journal of Econometrics* 165, 210-220.
- Koop, G. and Potter, S. 2010. A Flexible Approach to Parametric Inference in Nonlinear Time Series Models. *Journal of Econometrics* 159: 134-150.
- Korobilis, D. 2013. Assessing the Transmission of Monetary Policy Using Time-varying Parameter Dynamic Factor Models. *Oxford Bulletin of Economics and Statistics* 75: 157-179.
- Kulich, M., Morley, J., and Robinson, T., 2017. Estimating DSGE models with zero interest rate policy, *Journal of Monetary Economics* 88: 35-49.
- Lanne, M. and Luoto, J. 2017. A new time-varying parameter autoregressive model for US inflation expectations. *Journal of Money, Credit and Banking* 49(5): 969-995.
- Liu C., Rubin, D.B. and Wu, Y.N. 1998. Parameter Expansion to Accelerate EM: The PX-EM Algorithm. *Biometrika* 85, No. 4: 755-770

- Liu, J. and Wu, Y. 1999. Parameter expansion for data augmentation. *Journal of the American Statistical Association* 94: 1264-1274.
- Lubik, T. A. and Matthes, C. 2015a. Calculating the natural rate of interest: a comparison of two alternative approaches. *Economic Brief- Federal Reserve Bank of Richmond*, EB15-10.
- Lubik, T. A. and Matthes, C. 2015b. Time-varying parameter vector autoregressions: specification, estimation, and an application. *Economic Quarterly* Vol. 101, No. 4: 323-352.
- Mao, G. and Zhang, Z. 2018. Stochastic tail index model for high frequency financial data with Bayesian analysis. *Journal of Econometrics* 205: 470-487.
- McAlinn, K. and West, M. 2019. Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics*, 210: 155-169.  
<https://doi.org/10.1016/j.jeconom.2018.11.010>.
- McCausland, W., Miller, S., and Pelletier, D. 2011. Simulation smoothing for state-space models: A computational efficiency analysis, *Computational Statistics and Data Analysis* 55: 199-212.
- Morley, J. 2002. A state-space approach to calculating the Beveridge-Nelson decomposition. *Economics Letters* 75(1):123-127.
- Morley, J. 2010. The two interpretations of the Beveridge-Nelson decomposition. *Macroeconomic Dynamics* 15:419-439.
- Naesseth, C., Lindsten, F., and Schön, T. 2015. Nested sequential Monte Carlo methods. In *Proceedings of the 32nd International Conference on Machine Learning*, 1292–1301.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. 2007. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 140(2): 425-449.

Petrova, K. 2019. A quasi-Bayesian local likelihood approach to time varying parameter VAR models, *Journal of Econometrics* 212: 1: 286-306.

<https://doi.org/10.1016/j.jeconom.2019.04.031>.

Primiceri, G. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies*: 72: 821–852.

Raftery, A., Karny, M. and Ettlér, P. 2010. Online Prediction Under Model Uncertainty Via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics* 52, 52-66.

Rebeschini, P. and van Handel, R. 2015. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability* 25(5), 2809–2866.

Robert, S. and Künsch, H.R. 2017. Localizing the ensemble Kalman particle filter. *Tellus A: Dynamic Meteorology and Oceanography* 69(1), 1282016, DOI: 10.1080/16000870.2017.1282016.

Rue, H. 2001. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63: 325-338. doi:10.1111/1467-9868.00288

Sims, C. A., 2001, “Comment on Sargent and Cogley’s ‘Evolving Post World War II U.S. Inflation Dynamics,’” NBER Macroeconomics Annual 16, pp 373-379.

Smith, M. and Kohn, R. 1996. Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* 75: 317-344.

Stock, J.H. and Watson, M.W. 2007. Why has US inflation become harder to forecast?. *Journal of Money, Credit and banking*, 39: 3-33.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267-288.

Uribe, P., and Lopes, H. 2017. Dynamic sparsity on dynamic regression models. Manuscript, available at <http://hedibert.org/wp-content/uploads/2018/06/uribe-lopes-Sep2017.pdf>

Watson, M.W. 1986. Univariate Detrending with Stochastic Trends. *Journal of Monetary Economics* 18: 49-75.

Wigren, A., Murray, L., and Lindsten, F. 2018. Improving the particle filter in high dimensions using conjugate artificial process noise. In *18th IFAC Symposium on System Identification SYSID 2018 Proceedings* 51, 670–675.

DOI: 10.1016/j.ifacol.2018.09.207

Zhao, Z. Y., Xie, M and West, M. 2016. Dynamic dependence networks: Financial time series forecasting & portfolio decisions (with discussion). *Applied Stochastic Models in Business and Industry*, 32: 311-339.

<https://doi.org/10.1002/asmb.2161>.