# Time-Varying Parameter Vector Autoregressions: Specification, Estimation, and an Application

Thomas A. Lubik and Christian Matthes

T ime-varying parameter vector autoregressions (TVP-VARs) have become an increasingly popular tool for analyzing the behavior of macroeconomic time series. TVP-VARs differ from more standard fixed-coefficient VARs in that they allow for coefficients in an otherwise linear VAR model to vary over time following a specified law of motion. In addition, TVP-VARs often include stochastic volatility (SV), which allows for time variation in the variances of the error processes that affect the VAR.

The attractiveness of TVP-VARs is based on the recognition that many, if not most, macroeconomic time series exhibit some form of nonlinearity. For instance, the unemployment rate tends to rise much faster at the start of a recession than it declines at the onset of a recovery. Stock market indices exhibit occasional episodes where volatility, as measured by the variance of stock price movements, rises considerably. As a third example, many aggregate series show a distinct change in behavior in terms of their persistence and their volatility around the early 1980s when the Great Inflation of the 1970s turned into the Great Moderation, behavior that is akin to a structural shift in certain

moments of interest. All these examples of nonlinearity in macroeconomic time series have potentially distinct underlying structural causes. But they can all potentially be captured by means of the flexible framework that is a TVP-VAR with SV.

A VAR is a simple time series model that explains the joint evolution of economic variables through their own lags. A TVP-VAR preserves this structure but in addition models the coefficients as stochastic processes. In the most common application, the maintained assumption is that the coefficients follow random walks, specifically the intercepts, the lag coefficients as well as the variance and covariances of the error terms in the regression. Conditional on the parameters, a TVP-VAR is still a linear VAR, but the overall model is highly nonlinear. While the assumption of random walk behavior may seem restrictive, it provides for a flexible functional form to capture various forms of nonlinearity.

The main challenge in applying TVP-VAR models is how to conduct inference. In this article, we therefore discuss the Bayesian approach to estimating a TVP-VAR with SV.[1] Bayesian inference in this class of models relies on the Gibbs sampler, which is designed to easily compute multivariate densities. The key insight is to break up a computationally intractable problem into sequences of feasible steps. We will discuss these steps in detail and show how they can be applied to TVP-VARs.

The article is structured as follows. We begin with a discussion of the specification of TVP-VARs and how they are developed from fixed-coefficient VARs. We show how to introduce stochastic volatility in the covariance matrix of the errors and present an argument for why time variation in the lag coefficients needs to be modeled jointly with stochastic volatility. The main body of the article presents the Gibbs sampling approach to conducting inference in Bayesian TVP-VARs, which we preamble with a short discussion of the thinking behind Bayesian methods. Finally, we illustrate the method by means of a simple application to data on inflation, unemployment, and the nominal interest rate for the United States.

## 1.    SPECIFICATION

VARs are arguably the most important empirical tool for applied macroeconomists. They were introduced to the economics literature by Sims (1980) as a response to the then-prevailing large-scale macroeconometric modeling approach. What Sims memorably criticized were the

----

[1] Nakajima (2011) and Doh and Connolly (2012) provide similar overviews of the TVP-VAR methodology.

incredible identification assumptions imposed in these models that stemmed largely from a lack of sound theoretical economic underpinnings and that hampered structural interpretation of their findings. In contrast, VARs are deceptively simple in that they are designed to simply capture the joint dynamics of economic time series without imposing ad-hoc identification restrictions.

More specifically, a VAR describes the evolution of a vector of $n$ economic variables $y_t$ at time $t$ as a linear function of its own lags up to order $L$ and a vector $e_t$ of unforecastable disturbances:

$$y_t = c_t + \sum_{j=1}^{L} A_j y_{t-j} + e_t. \tag{1}$$

It is convenient to assume that the error term $e_t$ is Gaussian with mean 0 and covariance matrix $\Omega_e$. $c_t$ is a vector of deterministic components, possibly including time trends, while the $A_j$ are conformable matrices that capture lag dynamics.

VAR models along the lines of (1) have proven to be remarkably popular for studying, for instance, the effects and implementation of monetary policy (see Christiano, Eichenbaum, and Evans 1999, for a comprehensive survey). However, VARs of this kind can only describe economic behavior that is approximately linear and does not exhibit substantial variation over time. The linear VAR in (1) contains a built-in notion of time invariance: conditional forecasts as of time $t$, such as $E_t y_{t+1}$, only depend on the last $L$ values of the vector of observables but are otherwise independent of time. More strongly, the conditional one-step-ahead variance is fully independent of time: $E_t[(y_{t+1} - E_t y_{t+1})(y_{t+1} - E_t y_{t+1})'] = \Omega_e$.

Yet, in contrast, a long line of research documents that conditional higher moments can vary over time, starting with the seminal ARCH model of Engle (1982). Moreover, research in macroeconomics, such as Lubik and Schorfheide (2004), has shown that monetary policy rules can change over time and can therefore introduce nonlinearities, such as breaks or shifts, into aggregate economic time series.[2] The first observation has motivated Uhlig (1997) to introduce time variation in $\Omega_e$. The second observation stimulated the work by Cogley and Sargent (2002) to introduce time variation in $A_j$ and $c$ in addition to stochastic volatility.

---

[2] This feature would make a linear model less suited to capture the true dynamics of the economy. Whether and to what extent linear approximations can be used to analyze environments with time-varying parameters has been studied by Canova, Ferroni, and Matthes (2015).

We will now describe how to model time variation in each of these sets of parameters separately. In the next step, we will discuss why researchers should model changes in both sets of parameters jointly. We then present the Gibbs sampling algorithm that is used for Bayesian inference in this class of models and which allows for easy combination of the approaches because of its modular nature.

### A VAR with Random-Walk Time Variation in the Coefficients

Suppose a researcher wants to capture time variation in the data by using a parsimonious yet flexible model as in the VAR (1). The key question is how to model this time variation in the coefficients $A_j$ and $c$. One possibility is to impose a priori break points at specific dates. Alternatively, break points can be chosen endogenously as part of the estimation algorithm. Threshold VARs or VARs with Markov switching in the parameters (e.g., Sims and Zha 2006) are examples of this type of model, which is often useful in environments where the economic modeler may have some a priori information or beliefs about the underlying source of time variation, such as discrete changes in the behavior of the monetary authority. In general, however, a flexible framework with random time variation seems preferable for a wide range of nonlinear behavior in the data. Following Cogley and Sargent (2002), a substantial part of the literature has consequently opted for a flexible specification that can accommodate a large number of patterns of time variation.

The standard model of time variation in the coefficients starts with the VAR (1). In contrast to the fixed-coefficient version, the parameters of the intercept and of the lag coefficient matrix are allowed to vary over time in a prescribed manner. We thus specify the TVP-VAR:

$$y_t = c_t + \sum_{j=1}^{L} A_{j,t} \ y_{t-j} + e_t. \tag{2}$$

It is convenient to collect the values of the lagged variables in a matrix and define $X_t' \equiv I \otimes (1, y_{t-1}'..., y_{t-L}')$, where '$\otimes$' denotes the Kronecker product. We also define $\theta_t$ to collect the VAR's time-varying coefficients in vectorized form, that is, $\theta_t \equiv vec([c_t \ A_{1,t} \ A_{2,t} \ ... \ A_{L,t}]')$. This allows us to rewrite (2) in the following form:

$$y_t = X_t'\theta_t + e_t. \tag{3}$$

The commonly assumed law of motion for $\theta_t$ is a random walk:

$$\theta_t = \theta_{t-1} + u_t, \tag{4}$$

where $u_t \sim \mathcal{N}(0, Q)$ and is assumed to be independent of $e_t$. A random-walk specification is parsimonious in that it can capture a large number of patterns without introducing additional parameters that need to be estimated.[3] This assumption is mainly one of convenience for reasons of parsimony and flexibility as (4) is rarely interpreted as the underlying data-generating process for the question at hand, but it can approximate it arbitrarily well (see Canova, Ferroni, and Matthes 2015).

### Introducing Stochastic Volatility

A second source of time variation in time series can stem from variation in second or higher moments of the error terms. Stochastic volatility, or, specifically, time variation in variances and covariances, can be introduced into a model in a number of ways. Much of the recent literature on stochastic volatility in macroeconomics has chosen to follow the work of Kim, Shephard, and Chib (1998). It is built on a flexible model for volatility that uses an unobserved components approach.[4]

We start from the observation that we can always decompose a covariance matrix $\Omega_e$ as follows:

$$\Omega_e = \Lambda^{-1}\Sigma\Sigma' \left(\Lambda^{-1}\right)'. \tag{5}$$

$\Lambda$ is a lower triangular matrix with ones on the main diagonal, while $\Sigma$ is a diagonal matrix. Intuitively, the diagonal matrix $\Sigma\Sigma'$ collects the independent innovation variances, while the triangular matrix $\Lambda^{-1}$ collects the loadings of the innovations onto the VAR error term $e$, and thereby the covariation among the shocks. It has proven to be convenient to parameterize time variation in $\Omega_e$ directly by making the free elements of $\Lambda$ and $\Sigma$ vary over time. While this decomposition is general, once priors on the elements of $\Sigma$ and $\Lambda$ are imposed, the ordering of variables in the VAR matters for the estimation of the reduced-form parameters, which stands in contrast to the standard time-invariant VAR model (see Primiceri 2005).

We now define the element of $\Lambda_t$ in row $i$ and column $j$ as $\lambda_t^{ij}$ and a representative free element $j$ of the time-varying coefficient matrix $\Sigma_t$ as $\sigma_t^j$. It has become the convention in the literature to model the

---

[3] Different specifications for the time-varying lag coefficients are entirely plausible. For instance, a stationary VAR(1) representation, such as $\theta_t = \bar{\theta} + B\theta_{t-1} + u_t$, can easily be accommodated using the estimation algorithms described in this article.

[4] The approach to modeling stochastic volatility outlined here is the most common in the literature on TVP-VARs, but there are alternatives such as Rondina (2013). Moreover, stochastic volatility models of the form used here are more flexible than ARCH models in that they do not directly link the estimated level of the volatility to realizations of the error process that is being captured.

coefficients $\sigma_t^j$ as geometric random walks:

$$\log \sigma_t^j = \log \sigma_{t-1}^j + \eta_t^j. \tag{6}$$

For future reference, we collect the $\sigma_t^j$ in a vector $\sigma_t = \left[\sigma_t^1, ..., \sigma_t^n\right]'$ and the $\eta_t^j$ in $\eta_t = \left[\eta_t^1, ..., \eta_t^n\right]$, with $\eta_t^n \sim \mathcal{N}(0, W)$ and $W$ diagonal. Similarly, we assume that the nonzero and nonunity elements of the matrix $\Lambda_t$, which we collect in the vector $\lambda_t = \left[\lambda_t^{21}, ..., \lambda_t^{n,n-1}\right]$, evolve as random walks:

$$\lambda_t = \lambda_{t-1} + \zeta_t, \tag{7}$$

where $\zeta_t \sim \mathcal{N}(0, S)$ and $S$ block-diagonal.

The error term $e_t$ in the TVP-VAR representation (3) can thus be decomposed into:

$$e_t = \Lambda_t^{-1} \Sigma_t \varepsilon_t, \tag{8}$$

which implicitly defines $\varepsilon_t$. It is convenient to normalize the variance of $\varepsilon_t$ to unity. It is thus assumed that the error terms in each of the equations of the model are independent. In more compact form, we can write:

$$V = Var\left[\begin{pmatrix} \varepsilon_t \\ \zeta_t \\ \eta_t \end{pmatrix}\right] = \begin{bmatrix} I & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & W \end{bmatrix}. \tag{9}$$

The TVP-VAR literature tends to impose a block-diagonal structure for $V$, mainly for reasons of parsimony since the TVP-VAR is already quite heavily parameterized. Allowing for a fully generic correlation structure among different sources of uncertainty would also preclude any structural interpretation of the innovations. Following Primiceri (2005), the literature has therefore adopted a block-diagonal structure for $S$, which implies that the nonzero and non-one elements of $\Lambda_t$ that belong to different rows evolve independently. Moreover, this assumption simplifies inference substantially since it allows Kalman smoothing on the nonzero and non-one elements of $\Lambda_t$ equation by equation, as we will discuss further below.

### Why We Want to Model Time Variation in Volatilities *and* Parameters

A TVP-VAR with stochastic volatility is a heavily parameterized object. While it offers flexibility to capture a wide range of time variation and nonlinear features of the data, it also makes estimation and inference quite complicated. In practice, modelers restrict the covariance matrix of the innovations to the laws of motion for the time-varying

coefficients in order to sharpen inference. Moreover, Bayesian priors are often used to aid inference. Given a need to impose some structure to aid inference, this naturally raises the question whether a TVP-VAR with stochastic volatility is not overparameterized.

One answer to this question relies on the idea that a TVP-VAR can be regarded as the reduced-form representation of an underlying Dynamic Stochastic General Equilibrium (DSGE) model, in which there is time variation. This time variation in the underlying data-generating process (DGP) carries over to its reduced form, which might be, or is approximated by, a TVP-VAR.[5] More specifically, changes, discrete or continuous, in structural parameters carry over to changes in lagged reduced-form coefficients and parameters of the covariance matrix.[6] Hence, a TVP-VAR specification should a priori include stochastic volatility to be able to represent an underlying DSGE model.

A second response is essentially a corollary to the previous point. Sims (2002) argues that a model with only time variation in parameters could mistakenly result in a substantial amount of time variation even though the true DGP only features stochastic volatility. This insight can be illustrated by means of the following simple example, which also shows that the reverse can hold: a modeler could mistakenly estimate stochastic volatility even though the true DGP only features time variation in coefficients.

Consider a univariate AR(1)-process with stochastic volatility:

$$z_t = \rho z_{t-1} + \sigma_t \varepsilon_t, \tag{10}$$

where $|\rho| < 1$, $\varepsilon_t \sim \mathcal{N}(0, 1)$, and $\sigma_t$ is a generic stochastic volatility term, such as the one described above. Suppose an econometrician has access to a sample of data from this DGP, but does not know the true form of the underlying model. In order to investigate the time variation in the data, he proposes a model with only time-varying coefficients instead of stochastic volatility. As a simple rewriting of equation (10) suggests, he could indeed find evidence for time variation in the parameters:

$$z_t = \rho_t z_{t-1} + \tilde{\sigma} \varepsilon_t, \tag{11}$$

---

[5] It is well-known that in some cases a linear VAR is an exact representation of the reduced form of a DSGE model (see Fernandez-Villaverde et. al. 2007). It is less well-known to what extent this is true for TVP-VARs. For instance, Cogley, Sbordone, and Matthes (2015) show that DSGE models with learning have a TVP-VAR as reduced form.

[6] This insight underlies Benati and Surico's (2009) critique of Sims and Zha's (2006) Markov-switching VAR approach to identifying monetary policy shifts and also Lubik and Surico's (2010) critique of standard empirical tests of the validity of the Lucas critique.

where $\rho_t = \rho + \frac{(\sigma_t - \tilde{\sigma})\varepsilon_t}{z_{t-1}}$.

If the DGP is instead of the form:

$$z_t = \rho_t z_{t-1} + \sigma\varepsilon_t, \tag{12}$$

and estimates a stochastic volatility model on data generated from this model, he would erroneously find evidence of stochastic volatility:

$$z_t = \tilde{\rho} z_{t-1} + \sigma_t\varepsilon_t, \tag{13}$$

where $\sigma_t = \sigma + \frac{(\rho_t - \tilde{\rho})z_{t-1}}{\varepsilon_t}$. Including time variation jointly in coefficients and stochastic volatility therefore allows economists to let the data speak on which of the two sources are more important.

## 2.   ESTIMATION AND INFERENCE

A TVP-VAR with stochastic volatility is a deceptively simple object on the surface, as it superficially shares the structure of standard linear VARs. Estimation and inference in the latter case is well-established and straightforward. Since a linear VAR is a seemingly unrelated regression (SUR) model, it can be efficiently estimated equation by equation using ordinary least squares (OLS). Conducting inference on transformations of the original VAR coefficients, such as impulse response functions, is somewhat more involved yet well-understood in the literature. Estimation and inference in a TVP-VAR, however, reaches a different level of complexity since the model is fundamentally nonlinear due to the time variation in the coefficients and in the covariance matrix of the error terms.

We now describe in detail the standard approach to inference in TVP-VARs. It relies on Bayesian estimation, the basic concepts of which we introduce briefly in the following. Bayesian estimation and inference is conducted using the Gibbs sampling approach, which we go on to discuss at some length. Finally, we discuss how researchers can report and interpret the results from TVP-VAR models in a transparent and efficient manner.

### Why a Bayesian Approach?

The standard approach to estimating and conducting inference in TVP-VARs uses Bayesian methodology. The key advantage over frequentist methods is that it allows researchers to use powerful computational algorithms that are particularly well-adapted to the treatment of time variation. Moreover, the use of prior information in a Bayesian framework helps researchers to discipline the behavior of the model, which

is especially relevant in high-dimensional problems such as those discussed in this article.[7]

Bayesian and frequentist inference are fundamentally different approaches to describing and making assessments about data and empirical models. Bayesian inference starts by postulating a prior distribution for the parameters of the model. This prior is updated using the information contained in the data, which is extracted using a likelihood function. The object of interest in Bayesian estimation is the posterior distribution, which results from this updating process. Estimators in a Bayesian context are thus defined as statistics of this distribution such as the mean or mode.

We can describe these basic principles in a somewhat more compact and technical form. Suppose that a Bayesian econometrician is interested in characterizing his beliefs about parameters of interest $\Theta$ after having observed a sample of data $y^T$ of length $T$. The econometrician holds beliefs prior to observing the data, which can be described by the *prior* $p(\Theta)$. Moreover, he can summarize the data by computing the *likelihood function* $p(y^T|\Theta)$, which describes how likely the observed data are for any possible parameter vector $\Theta$. The beliefs held by the econometrician after seeing the data are summarized by the *posterior distribution* $p(\Theta|y^T)$. The relationship between those three densities is given by *Bayes' law*:

$$p(\Theta|y^T) = \frac{p(y^T|\Theta)p(\Theta)}{\int p(y^T|\Theta)p(\Theta)d\Theta}, \tag{14}$$

which describes how to optimally update the beliefs contained in $p(\Theta)$ using data summarized by $p(y^T|\Theta)$. The posterior $p(\Theta|y^T)$ is a distribution on account of normalization by the *marginal data density* $\int p(y^T|\Theta)p(\Theta)d\Theta$, which is the joint distribution of data $y^T$ and parameters $\Theta$ after integrating out $\Theta$. It can serve as a measure of fit in this Bayesian context.

Bayesian estimation ultimately consists of computing the posterior distribution. Bayesian inference rests on the moments of this distribution. It does not require any arguments about limiting behavior as $T \to \infty$, since from a Bayesian perspective $y^T$ is fixed and is all that is needed to conduct inference. On the other hand, the challenges for Bayesian econometricians are virtually all computational in that: (i) the likelihood function has to be evaluated; (ii) the joint distribution of prior and likelihood has to be computed; and (somewhat less

---

[7] This is not to say that frequentist inference does not introduce prior information by, for instance, imposing bounds on the parameter space. The use of Bayesian priors, however, makes this more explicit and generally more transparent.

crucially) (iii) the marginal data density has to be obtained. What aids in this process is the judicious use of priors and fast and robust methods for characterizing $p(y^T|\Theta)p(\Theta)$. This can be accomplished in Bayesian VARs by means of the Gibbs sampler.

### Gibbs Sampling of a TVP-VAR

Characterizing a posterior distribution is a daunting task. Except in special cases, analytical solutions for given prior and likelihood densities are not available. Conducting inference via describing the posterior with its moments is thus not an option. As evidenced by the seminal textbook of Zellner (1971), much of Bayesian analysis before the advent of readily available computing power and techniques was concerned with finding conjugate priors for a large variety of problems. A conjugate prior is such that when confronted with the likelihood function, the posterior distribution is of the same family as the prior. However, as a general matter this path proved not to be a viable option as many standard Bayesian econometric models do not easily yield to analytical characterization.

This changed with the development of sampling and simulation methods that allow researchers to characterize the shape of an unknown distribution. These methods are built on the idea that when a large sample from a known density is available, sample moments approximate population moments very well by the laws of large numbers. Consequently, Bayesian statisticians have developed methods to efficiently sample from unknown posterior densities indirectly by sampling from known densities. Once the thus-generated sample is at hand, sampling moments can be used to characterize the posterior distribution.[8]

The basic idea behind the Gibbs sampler is to split the parameters $\Theta$ of a given model into $b$ blocks $\Theta^1, \Theta^2, ..., \Theta^b$.[9] The Gibbs sampler proposes to generate a sample from $p(\Theta|y^T)$ by iteratively sampling from $p(\Theta^j|y^T, \Theta^{-j})$, $\forall j = 1, ..., b$, where $\Theta^{-j}$ denotes the entire parameter vector except for the *jth* block. This approach rests on the idea that the entire set of conditional distributions fully characterizes the joint distribution under fairly general conditions. At first glance, nothing

---

[8] The exposition here is intentionally, but unavoidably, superficial. Readers interested in the technical issues underlying the arguments we make here are referred to some of the excellent textbooks on Bayesian inference such as Robert and Casella (2004) or Gelman et al. (2014).

[9] Generally, there are no restrictions placed on the relative size of the blocks. In fact, the blocking scheme, that is, its individual size, could be random. However, for time-varying parameter models, one particular blocking scheme turns out to be especially useful.

much has been gained: we have broken up one large inference problem into a sequence of smaller inference problems, namely characterizing the conditional distributions $p(\Theta^j | y^T, \Theta^{-j})$ instead of the full distribution. In the end, there is no guarantee that this makes the inference problem more tractable.

However, Bayesian statisticians have developed closed forms for posterior distributions for some special cases. The ingenuity of the Gibbs sampler is thus to break up a large intractable inference problem into smaller blocks that can then be evaluated independently and sequentially. The challenge is to find a blocking scheme, a partition of the set of parameters, that admits closed-form solutions for the posteriors conditional on all other parameters of the model. In the case of TVP-VARs, such blocking schemes have been developed by Cogley and Sargent (2002), Primiceri (2005), and Del Negro and Primiceri (2015).[10]

### A Motivating Example for the Gibbs Sampler

In order to illustrate the basic idea behind Gibbs sampling, we consider a simple fixed-coefficient AR(1) model:

$$z_t = \rho z_{t-1} + \sigma \varepsilon_t, \tag{15}$$

where $\varepsilon_t \sim \mathcal{N}(0, 1)$. The parameters of interest are $\rho$ and $\sigma$, on which we want to conduct inference. The first step in deriving the Gibbs sampler is to specify priors for these parameters. We assume the following priors:

$$\rho \sim \mathcal{N}(\mu_\rho, V_\rho), \tag{16}$$

$$\sigma^2 \sim \mathcal{IG}(a, b), \tag{17}$$

where $IG$ denotes the inverse Gamma distribution with scale and location parameters $a$ and $b$, respectively.

The likelihood for this standard AR(1) model is given by $L(\rho, \sigma) = p(z_0) \prod_{t=1^T} p(z_t | z_{t-1})$, which is written as the product of conditional distributions $p(z_t | z_{t-1})$ and the likelihood of the initial observation $p(z_0)$. As is common practice, we drop the term $p(z_0)$ and instead work with the likelihood function $L(\rho, \sigma) = \prod_{t=1}^{T} p(z_t | z_{t-1})$. Defining $Y = [z_1 \ z_2 \ z_3 \ldots z_T]'$ and $X = [z_0 \ z_1 \ z_2 \ldots z_{T-1}]'$, the likelihood is given by:

$$L(\rho, \sigma) = (2\pi)^{-T/2} (\sigma^2)^{-T/2} \exp\left[ -\frac{1}{2\sigma^2} (Y - X\rho)'(Y - X\rho) \right]. \tag{18}$$

---

[10] Computer code to estimate this class of models is available from Gary Koop and Dimitris Korobilis at: http://personal.strath.ac.uk/gary.koop/bayes_matlab_code_by_koop_and_korobilis.html

Combining this expression with the priors listed above using Bayes' Law gives the joint posterior of $\rho$ and $\sigma^2$, conditional on the data:

$$p(\rho, \sigma | Y, X) \propto L(\rho, \sigma) \times$$

$$\exp\left[-\frac{1}{2}(\rho - \mu_\rho)' V_\rho^{-1} (\rho - \mu_\rho)\right] \times (\sigma^2)^{-[a+1]} \exp\left(-\frac{1}{b\sigma^2}\right), \quad (19)$$

where the first term is the likelihood function, the second is the prior on the autoregressive coefficient $\rho$, and the third term is the prior on the innovation variance $\sigma^2$. Although we can identify and compute analytically the individual components of the posterior, the posterior distribution for $\rho, \sigma | Y, X$ is unknown.

The Gibbs sampler allows us to partition the parameter set into separate blocks for $\rho$ and $\sigma$, for which we can derive the conditional distributions. After some algebra, we can find the conditional posterior distributions:

$$\rho | \sigma, Y, X \sim \mathcal{N}\left[\begin{array}{c} \left(X'X/\sigma^2 + V_\rho^{-1}\right)^{-1} \left(X'Y/\sigma^2 + V_\rho^{-1}\mu_\rho\right), \\ \left(X'X/\sigma^2 + V_\rho^{-1}\right)^{-1} \end{array}\right], \quad (20)$$

$$\sigma^2 | \rho, Y, X \sim \mathcal{IG}\left[T/2 + a, b^{-1} + \frac{1}{2}(Y - X\rho)'(Y - X\rho)\right]. \quad (21)$$

The conditional posteriors for $\rho$ and $\sigma$ have known distributions, which can be sampled by using standard software packages. The procedure would be to start with an initial value for $\sigma^2$ and then draw from the conditional distribution $\rho | \sigma, Y, X$. Given a draw for $\rho$, in the next step we would sample from the conditional distribution $\sigma^2 | \rho, Y, X$. Repeated iterative sampling in this manner results in the joint posterior distribution $\rho, \sigma | Y, X$.

The Gibbs sampler can be applied to models with time-varying parameters in a similar manner, the key step being the application of a blocking scheme for which the conditional distributions are either known or from which it is easy to generate samples. The additional challenge that TVP-VARs present is that the parameters of interest are not fixed coefficients, but are themselves time-series processes that are a priori unobservable. The general approach to dealing with unobservable components is the application of the Kalman filter if the model can be cast in a state-space form. In the following, we discuss how these two additional techniques can be used to estimate TVP-VARs.

### *Linear Gaussian State-Space Systems*

Bayesian estimation relies on the ability of the researcher to cast a model in a form such that it is amenable for sampling. The Gibbs

sampler provides one such technique. A second crucial component of inference in a TVP-VAR is the state-space representation, which connects variables that are observed, or are in principle observable, to those that are unobserved. Conceptually, Bayesian estimation produces a time series and its density for the time-varying components of the TVP-VAR by means of the Kalman filter as applied to a linear Gaussian state-space system. This specification has the advantage that the posterior distribution is known analytically for a Gaussian prior on the initial state.

Specifically, a state-space system can be defined as follows:

$$y_t = A_t x_t + B_t v_t, \tag{22}$$
$$x_t = C x_{t-1} + D w_t, \tag{23}$$

where $y_t$ denotes a vector of observables and $x_t$ a vector of possibly unobserved states. $v_t$ and $w_t$ are Gaussian innovations, each element of which is independent of the others with mean 0 and variance 1. $A_t$, $B_t$, $C$, and $D$ are *known* conformable matrices. The standard approach for deriving the posterior for $x_t$ in this system was developed by Carter and Kohn (1994), which builds on the Kalman filter and which we discuss in the next section.

Application of the Kalman filter to a state-space system allows the modeler to construct a sequence of Gaussian distributions for $x_t | y^t$, that is, the distribution of the unobservable state $x$ at time $t$, conditional on the observables $y^t$, where a superscript denotes the entire sample up to that point.[11] As it turns out, various blocks of the Gibbs sampler for a TVP-VAR model take the form of linear Gaussian state-space systems. The challenge is to find blocks for the parameters in the TVP-VAR such that each block fits this Gaussian state-space structure. The fundamental nonlinearity of the TVP-VAR can thus be broken up into parts that are conditionally linear and from which it can be easily sampled. As long as each block has a tractable structure conditional on other blocks of parameters, the Gibbs sampler can handle highly nonlinear problems.

### The Kalman Filter

The Kalman filter is a widely used method for computing the time paths of unobserved variables from a Gaussian state-space system. We now briefly review and present the equations used for drawing a sequence of the unobserved states (conditional on the entire set of observations

---

[11] If the modeler is instead interested in the distributions $x_t | y^T$, where $T$ denotes the sample size, the Carter-Kohn algorithm draws paths of the unobserved state variable $x_t$ for $t = 1, ..., T$ conditional on the entire sample of observables $y^T$.

$y_1, ..., y_T$). A more detailed discussion and explanation can be found in Primiceri (2005).

The system is assumed to take the form (22)-(23). We want to draw from the distribution $p(x_1, ..., x_T | y_1, ..., y_T)$.[12]  It can be shown that $p(x_1, ..., x_T | y_1, ..., y_T) = p(x_T | y_T) \prod_{t=1}^{T} p(x_t | x_{t+1}, y_1, ..., y_t)$. To generate draws from each of these densities, we first run the Kalman filter to calculate the mean and variance of the state $x_t$ conditional on data up to time $t$. We assume a prior for $x_0$ that is Gaussian with mean $x_{0|0}$ and variance $V_{0|0}$. The Kalman filter is then summarized by the following equations:

$$x_{t|t-1} = Cx_{t-1|t-1} \tag{24}$$

$$V_{t|t-1} = CV_{t-1|t-1}C' + DD' \tag{25}$$

$$K_t = V_{t|t-1}A_t' \left( A_t V_{t|t-1} A_t' + B_t B_t'^{-1} \right) \tag{26}$$

$$x_{t|t} = x_{t|t-1} + K_t(y_t - A_t x_{t|t-1}) \tag{27}$$

$$V_{t|t} = V_{t|t-1} - K_t A_t V_{t|t-1} \tag{28}$$

These equations produce $x_{t|t} = E(x_t | y_1, ..., y_t)$ and the associated conditional variance $V_{t|t}$. The conditional distributions of the states are Gaussian.

We can generate a draw for $x_T | y_1, ..., y_T$ by using the conditional mean and variance for period $T$. Once we have such a draw, we can recursively draw the other states ($x_{t+1}$ denotes a draw of the state for period $t + 1$):

$$x_{t|t+1} = x_{t|t} + V_{t|t}CV_{t+1|t}^{-1}(x_{t+1} - Cx_{t|t}) \tag{29}$$

$$V_{t|t+1} = V_{t|t} - V_{t|t}C'V_{t+1|t}^{-1}CV_{t|t} \tag{30}$$

In the following, we will now discuss each Gibbs sampler step in turn, which builds on the Kalman filter.

### The Choice of Priors

The first step in Bayesian analysis is to choose the priors on the parameters of the model. In contrast to a frequentist approach, the model parameters in a Bayesian setting are random variables. Since a Gibbs sampler proceeds iteratively, we impose priors on the initial values of the TVP-VAR parameters. Conceptually, it is therefore useful to distinguish between two sets of parameters: the parameters associated with the coefficients and innovation terms in the representation (4)

---

[12] We do not explicitly state the dependence of the densities in this section on the system matrices $A$, $B$, $C_t$, and $D_t$, but as we show later this can be handled by the right conditioning and sequencing within the Gibbs sampler.

and the parameters governing the law of motion of the time-varying terms. More specifically, we impose priors on $(\theta_0, \Lambda_0, \log \Sigma_0)$ and on $(Q, W, S)$, respectively.

The initial values of the lag coefficient matrices $\theta_0$, of the free elements of the loading matrix in the innovation terms $\Lambda_0$, and of the independent innovation variances $\log \Sigma_0$ are assumed to have normally distributed priors:

$$\theta_0 \sim \mathcal{N}(\overline{\theta}, \kappa_\theta V_\theta), \tag{31}$$

$$\Lambda_0 \sim \mathcal{N}(\overline{\Lambda}, \kappa_\Lambda V_\Lambda), \tag{32}$$

$$\log \Sigma_0 \sim \mathcal{N}(\overline{\Sigma}, I), \tag{33}$$

where $\overline{\theta}$, $\overline{\Lambda}$, and $\overline{\Sigma}$ are the prior means of the respective variables, while $V_\theta$ and $V_\Lambda$ are their prior covariance matrices. The covariance matrix of the prior on $\log \Sigma_0$ is normalized at unity. $\kappa_\theta$ and $\kappa_\Lambda$ are scaling parameters that determine the tightness of the priors.

We also have to choose priors for the covariance matrices of the innovations in the law of motions for the above-referenced TVP-VAR parameters. These are, respectively, the innovation variance for the lag coefficient matrices, $Q$; for the error variance, $W$; and for the loading matrix, $S$. As is common for covariance matrices in Bayesian analysis, the priors follow an Inverted Wishart distribution:

$$Q \quad \sim \quad \mathcal{IW}(\kappa_Q^2 df_Q V_Q, df_Q), \tag{34}$$

$$W \quad \sim \quad \mathcal{IW}(\kappa_W^2 df_W V_W, df_W), \tag{35}$$

$$S \quad \sim \quad \mathcal{IW}(\kappa_S^2 df_S V_S, df_S), \tag{36}$$

where $\kappa$ are the scaling factors, $df$ the degrees of freedom, and the matrices $V$ the respective variances.

A key issue is how to choose the parameters for the priors. Cogley and Sargent (2005) and Primiceri (2005) propose using a constant-coefficient VAR estimated on a training sample to initialize the prior means and the matrices $V$. The coefficients $(\overline{\theta}, \overline{\Lambda}, \overline{\Sigma})$ and $(V_\theta, V_\Lambda)$ can then be directly computed from a least-squares regression. Nevertheless, this still leaves substantial degrees of freedom as there is no clear guideline on how to choose the training sample. The scaling parameters $\kappa$ turn out to be important as they govern the prior amount of time variation. Primiceri (2005) estimates the $\kappa$ on a small grid of values using a time-consuming reversible-jump MCMC algorithm that, as a preliminary step, requires estimation of the model for each possible combination of parameters. Following Primiceri, most researchers

have chosen to use his estimated values regardless of the application at hand.[13]

### The Ordering of Blocks in a TVP-VAR

Once the priors have been chosen, the next step involves combining the prior distribution with the likelihood function. In a Bayesian approach, the resulting posterior distribution contains all information that is available to the researcher, which includes the prior and the observed data as encapsulated in the likelihood. Moreover, and in contrast to a frequentist approach to inference, Bayesian estimation does not involve an actual estimation step, where an estimator, that is, a mapping from data to the object of interest that satisfies some desirable criteria, is derived. Bayesian estimation simply involves characterizing the posterior distribution, which can be accomplished in the case of a TVP-VAR by means of the Gibbs sampler. A Bayesian econometrician then finds it often convenient to report moments of the posterior as estimation results.

The Gibbs sampler relies on the idea that it is often much easier to sequentially sample from conditional distributions, whose probability laws may be known, than from an unknown distribution. The tricky and often difficult part of this approach is to partition the parameter space into blocks such that this sampling is feasible and can be accomplished efficiently. To wit, in the full TVP-VAR model with both time-varying parameters and stochastic volatility, we need to estimate the following set of parameters: $\theta^T$, $\Sigma^T$, $\Lambda^T$, $Q$, $S$, and $W$, where the $T$ superscripts indicate that there can be in general sample size $T$ parameters.

In the following, we describe the Gibbs sampler proposed by Del Negro and Primiceri (2015), which is based on the original contribution of Primiceri (2005). As a matter of notation, we also introduce a set of auxiliary variables $s^T$ that are used for the estimation of the stochastic volatilities. In subsequent sections we discuss the drawing of each of those blocks in more detail. Even more detailed descriptions can be found in Primiceri (2005) or Koop and Korobilis (2010).

Conceptually, the two main steps of the Gibbs sampler involve drawing the covariance matrix of the independent innovations in the TVP-VAR, $\Sigma^T$, conditional on the data, the other coefficient vectors,

---

[13] In the recent literature, there has been much interest in the role that these scaling parameters play, in particular the hyperparameters for $Q$, $S$, and $W$. As it turns out, choice of these parameters can affect estimation results along many dimensions. For a recent application that studies the importance of these hyperparameters in producing the 'correct' inference see Lubik, Matthes, and Owens (2016).

and the covariance matrices of the processes governing time variation. In the second step, the remaining parameters are drawn from a distribution conditional on the data and on the draw from the first step $\Sigma^T$. Specifically, the procedure is to

1. draw $\Sigma^T$ from $p(\Sigma^T|y^T, Q, S, W, \Lambda^T, \theta^T, s^T)$

2. draw $\Lambda^T$, $\theta^T$, $s^T$, $Q$, $S$, and $W$ from $p(Q, S, W, \Lambda^T, \theta^T, s^T|y^T, \Sigma^T)$.

The second step is implemented as a sequence of intermediate steps. First, the algorithm draws from $p(Q, S, W, \Lambda^T, \theta^T|y^T, \Sigma^T)$, while the auxiliary variables $s^T$ are then drawn from $p(s^T|Q, S, W, \Lambda^T, \theta^T, y^T, \Sigma^T)$. This second step is split up into these two parts since this blocking scheme allows drawing $\theta^T$ without having to condition on $s^T$. Specifically, the sequence is to

i) draw $\Lambda^T$ from $p(\Lambda^T|y^T, \Sigma^T, Q, S, W, \theta^T)$

ii) draw $Q$, $S$ and $W$ from $p(Q, S, W|y^T, \Lambda^T, \Sigma^T, \theta^T)$

iii) draw $\theta^T$ from $p(\theta^T|y^T, Q, S, W, \Lambda^T, \Sigma^T)$

iv) draw $s^T$ from $p(s^T|y^T, \theta^T, Q, S, W, \Lambda^T, \Sigma^T)$.

### Drawing $\Sigma^T$

The first step of the Gibbs sampler involves generating draws of the elements of covariance matrix $\Sigma^T$ from a distribution that is conditional on the data $y^T$ and the remaining coefficient matrices. This conditional distribution conflates elements of the prior and the likelihood function; it is, in fact, a marginal density of the posterior. Draws are realizations of the random variable $\Sigma^T$ and are accordingly recorded. We now describe how a known conditional probability distribution for $\Sigma^T$ can be derived under this blocking scheme.

We can rewrite equation (3) under the assumption that $e_t$ features stochastic volatility:

$$\Lambda_t \left( y_t - X_t'\theta_t \right) = y_t^* = \Sigma_t \varepsilon_t, \tag{37}$$

where we have made use of the decomposition of the errors in equation (8). Given the conditioning set of this block in Step 1 above, $y_t^*$ is known. We can nowcast this representation into a Gaussian state-space system to draw the elements of $\Sigma^T$. Squaring each element of this vector and taking natural logarithms yields for each element $i$ of

$y_t^*$:[14]

$$\log\left((y_{i,t}^*)^2\right) = y_{i,t}^{**}. \tag{38}$$

We define $\sigma_t$ as the vector of the diagonal elements of $\Sigma_t$.

We then get the state-space system:

$$
\begin{aligned}
y_t^{**} &= 2\log(\sigma_t) + 2\log(\varepsilon_t), & (39)\\
\log(\sigma_t) &= \log(\sigma_{t-1}) + \eta_t. & (40)
\end{aligned}
$$

This is a linear state-space system with $y_t^{**}$ being the observable variable, while $\log(\sigma_t)$ is the unobserved state variable. However, it is not Gaussian: each element of $2\log(\varepsilon_t)$ is distributed as $\log\chi^2$ since it is the log of the square of a standard-normal random variable. These shocks can be approximated with a mixture of seven normal variables, as suggested by Kim, Shephard, and Chib (1998). In this step, the auxiliary variables $s^T$ are introduced to provide a record of which of the seven mixture components is 'active' for each element of $2\log(\varepsilon_t)$. Given this approximation, we have another Gaussian state-space system, which can now be evaluated using the Kalman filter. The prediction formulas listed above can be used to generate realizations, that is, draws, of the unobservable $\sigma_t$ over time.

### Drawing $\Lambda^T$

Given the draws for the matrix $\Sigma^T$, which is a component of the reduced-form error matrix $\Omega_{e,t}$ per equation (8), we can now sample its other component, namely the loadings $\Lambda^T$. The first step is to rewrite equation (3) but utilizing a different blocking:

$$\Lambda_t(y_t - X_t'\theta_t) = \Lambda_t\hat{y}_t = \Sigma_t\varepsilon_t. \tag{41}$$

The difference to the previous sampling scheme for $\Sigma^T$ is that we now condition on $\Sigma^T$ and are interested in sampling the free elements of the lower-triangular matrix $\Lambda_t$.

We can therefore rewrite the equation above by moving elements of $\Lambda_t\hat{y}_t$ to the right-hand side. We can write:

$$\hat{y}_t = Z_t\lambda_t + \Sigma_t\varepsilon_t, \tag{42}$$

where $Z_t$ is a selection matrix that contains elements of the vector $\hat{y}_t$. Together with the set of equations (7), this equation forms a linear Gaussian state-space system. The fact that $Z_t$ depends on elements of $\hat{y}_t$ poses no problem for the sampling step under the assumption

---

[14] In practice, and in order to improve numerical stability, we instead define $\log((y_{i,t}^*)^2 + c) = y_{i,t}^{**}$, where $c$ is a small 'offset' constant.

that the innovation covariance matrix for $\lambda$, $S$, is block diagonal. The Kalman filter can then be used to obtain draws for $\Lambda^T$.

### Drawing Innovation Covariance Matrices

In the next step, we are drawing from the innovation covariance matrices for the processes governing the time variation of the VAR parameters. As discussed above, each of the matrices $Q$, $S$, and $W$ is assumed to have an inverse-Wishart prior to facilitate the application of the Kalman filter within a Gaussian state-space system. In combination with a normally distributed likelihood, this prior forms a conjugate family since the innovations in the laws of motion for parameters and volatilities are Gaussian. Consequently, the posterior will also be of the inverse-Wishart form, which has a closed-form representation.[15] It is then straightforward to sample the innovation covariance matrices by drawing from the known inverted-Wishart posterior.

### Drawing $\theta^T$

In a penultimate step, we are now ready to sample from the conditional distribution for the TVP-VAR coefficient matrices. Given the preliminary work up this point and the use of the conditioning scheme that we describe above, this is now straightforward. Since we condition on draws for the covariance matrix of $e_t$, which in the general model with stochastic volatility will consist of draws for $\Lambda_t$ and $\Sigma_t$, equations (3) and (4) form a Gaussian state-space system. We can sample from the posterior distribution for $\theta^T$ in the manner described above by using the Kalman prediction equations to sequentially construct the draws.

### Drawing $s^T$

The final step that brings everything together involves the auxiliary variables $s^T$ that we use to track the stochastic volatilities. As we discuss above, each element of $s_t$ is drawn from a discrete distribution, a mixture of normals, with seven possible outcomes. Denote the prior probability for outcome $j$ as $q_j$. The conditional posterior probability used to drawing outcome $j$ for each element of $s^T$ is then proportional to

$$q_j f_N(y_{i_t}^{**}, 2\log(\sigma_{i,t}) + m_j, v_j^2), \tag{43}$$

---

[15] See, for example, Gelman et al. (2014).

where $m_j$ and $v_j$ are the given mean and standard deviation of each element of the Gaussian approximation and $f_N(x, a, b)$ is the Gaussian density with argument $x$, mean $a$, and variance $b$.

## Reporting the Results

Estimating a Bayesian TVP-VAR is tantamount to sampling from a posterior distribution. While the posterior summarizes all information available in the data and in the prior, it is an unwieldy object in that it is a multivariate distribution of which only the conditional distributions are known. The Gibbs sampling algorithm solves this problem by sequentially building up the joint distribution from the conditional distributions. Yet, what Bayesian estimation delivers are distributions and not point estimates. Reporting the results in a manner that is useful for economic interpretation therefore requires some thought. The Bayesian literature focuses on posterior means or medians as counterparts to frequentist point estimates. Instead of standard errors and confidence intervals, Bayesians report coverage regions that essentially are regions of the posterior distribution in which a given percentage of draws fall around a focal point such as the mean or the median.

The results from Bayesian fixed-coefficient VARs can be reported in a similar manner as for frequentist approaches. The reporting problem is compounded, however, in the case of TVP-VARs, since the distribution of the VAR parameters potentially changes at every data point, which is the very definition of time variation. Instead of reporting a single distribution in the case of a fixed-coefficient VAR, the Bayesian econometrician now faces the challenge of reporting a sequence of distributions. We describe in the following how to approach this issue for the case of impulse response functions, which are key objects in the toolkits of time series econometricians.

### *Impulse Responses*

VARs can be used to study the effects of exogenous shocks, that is, of unpredictable changes in the economy. For this purpose, the main tool in VAR analysis is the impulse response function that describes the behavior of a variable in response to a shock over time. In order to understand the sources of business cycles or to analyze policy, it is often desirable to give these shocks a structural interpretation. By doing so, researchers can link the shocks to economic theories.[16] However, the

---

[16] In line with time-invariant VARs, the literature usually focuses on studying the effects of shocks to observables, not shocks to the parameters that vary over time.

shocks that are estimated as residuals from a regression of the type (1) are generally not useful for this purpose as they conflate the effects of underlying structural disturbances. That is, the estimated residuals are generally correlated, in which case it is not possible to identify the effects of an individual disturbance.

More specifically, a researcher may be interested in uncovering uncorrelated disturbances $w_t$ that are a linear function of the regression errors $e_t$:

$$H_t w_t = e_t, \tag{44}$$

where it is assumed that $w_t$ is Gaussian with mean zero and a covariance matrix that is normalized to unity, $w_t \sim \mathcal{N}(0, I)$. The conformable matrix $H_t$ is used to transform the errors $e_t$ into the structural shocks $w_t$. How to derive and impose restrictions on $H_t$ is one of the key issues in VAR analysis. For instance, the economic theories used to define the shocks, e.g., DSGE models, can be used to derive restrictions on $H_t$. For the most part, it is common practice in the VAR literature to focus on imposing few enough restrictions so that the restrictions do not alter the likelihood function of the model. This has the advantage that the researcher can first estimate a statistical, 'reduced-form' model without worrying about the restrictions used to derive structural shocks. Structural shocks can then be studied after the estimation step is completed.[17]

For purposes of exposition we now discuss the most common and straightforward method for identifying structural shocks. It only assumes restrictions on the within-period-timing of shocks. The specific idea is that some shocks may be causally prior to other shocks in the sense that they have an impact on some variables and not on others within the period. The easiest way to implement this restriction is to make $H_t$ lower triangular. This can be achieved by calculating the Cholesky decomposition of the covariance matrix of the forecast errors.

In the context of TVP-VARs, this type of recursive ordering is appealing because $\Lambda_t^{-1} \Sigma_t$ already has lower triangular form so that the matrix $H_t$ can be directly calculated from the output of the Gibbs sampler. Given $H_t$, the impulse responses can then be calculated by simulation.[18] In contrast to fixed-coefficient VARs, it is thus not

---

[17] Using more restrictions so that the likelihood function is altered relative to the estimation of a reduced-form model means that the restrictions have to be imposed during estimation, that is, a 'structural model' has to be estimated directly. This is not often carried out, even though algorithms are now available even in the context of TVP-VARs, for instance in Canova and Perez-Forero (2015).

[18] A simpler method to approximate impulse responses is to draw a set of parameters from the Gibbs sampler output for each time period $t$ and then compute im-

possible to separate the estimation from the identification stage. In this case, the estimated variance-covariance matrix can be decomposed into its recursive components after the VAR is estimated. A detailed description of the algorithm is available in Canova and Gambetti (2009). We briefly describe the algorithm below.

Conceptually, we can define an impulse response as the difference between the expected path of the variables in the model when a shock of a given size hits and the expected path of the same variables when all shocks are drawn randomly from their distributions. In order to calculate impulse responses starting at time $t$, the first step is to draw a set of parameters from the Gibbs sampling output. Next, paths of future time-varying parameters and volatilities and a sequence of $w$ shocks are simulated once the identification matrix $H_t$ is computed. These objects are then used to calculate one path for the variables of interest using equation (2). The same exercise is repeated, but with the value of one structural shock fixed at one point in time, leaving all other structural shocks at the simulated values. This yields another path for the variables of interest, so that the difference between the paths is one realization of the impulse response. This sequence is repeated a large number of times for different parameter draws from the posterior and different simulated values of parameter paths and shocks. The approach produces a distribution of a path for the impulse responses for each time period in the sample. To report the results, the literature usually either picks a subset of time periods and then plots the median response as well as posterior bands for each time period separately or authors focus on the posterior median responses and plot those over time and for different horizons in a three-dimensional plot.[19]

## 3.    APPLICATION: A SIMPLE TVP-VAR MODEL FOR THE UNITED STATES

We now apply the methods discussed above to three key economic variables: the inflation rate, the unemployment rate, and a nominal interest rate. These three variables form the core of many models that are used to analyze the effects of monetary policy, such as the standard New Keynesian framework. Moreover, they are staples in most VARs that are used for the analysis of monetary policy. In his seminal paper,

---

pulse responses as if those parameters at time $t$ were parameters of a fixed coefficient VAR. This approach is computationally easier but neglects the fact that parameters and volatilities can change in the future.

[19] An example of the former can be found in Benati and Lubik (2014), while the latter approach is used in Amir-Ahmadi, Matthes, and Wang (2016).

Primiceri (2005) estimates a TVP-VAR in these three variables to study the effects of monetary policy in the post-World War II period in the United States. We base our application on his specification.

We update the data set to include more recent observations. The full sample ranges from the first quarter of 1953 to the first quarter of 2007, before the onset of the Great Recession. The data are collected quarterly, whereby percentage changes are computed on a year-over-year basis. As our measure of inflation, we use the (log-difference of the) GDP deflator, reported in percentage terms. As our economic activity variable, we pick the headline unemployment rate, while we use the three-month Treasury bill rate as the nominal interest rate variable. The data series are extracted from the FRED database at the Federal Reserve Bank of St. Louis.

We follow Primiceri (2005) in selecting a lag length of two for the TVP-VAR. This choice has become common in the TVP-VAR literature. In fixed-coefficient VARs, a higher number of lags is usually used, but the higher degree of complexity and dimensionality imposes nontrivial computational constraints. A lag length of two thus seems a good compromise and also allows for direct comparison of our results with other key papers in the literature. As discussed above, we need to provide an initialization for the prior. We follow common practice and use the first ten years of data for this purpose. The remaining priors are as in Primiceri (2005).

The first set of results that we extract from our TVP-VAR is contained in Figure 1. We report the median coefficient estimates from our model in three separate panels. The plots start with the first quarter of 1963 because the first ten years of the sample were used for the initialization of the prior. The upper panel contains plots of the time-varying lag coefficients $A_{j,t}$ and the intercept $c_t$ from equation (2). The overriding impression is that there is not much time variation in the lag coefficients. This is a finding that occurs throughout much of the TVP-VAR literature. However, evidence of some more time variation is apparent from the middle and lower panels, which report the time-varying components of the reduced-form innovation variance $\Omega_{e,t} = \Lambda_t^{-1} \Sigma_t \Sigma_t' \left( \Lambda_t^{-1} \right)'$.

The middle panel contains the nonzero and nonunity elements of the lower triangular matrix $\Lambda_t^{-1}$. The three off-diagonal elements are thus related to the correlation pattern in the estimated covariance matrix of the shocks. The panel shows that the relationship between inflation and the interest rate errors is consistently negative throughout the sample, while it is positive between the interest rate and unemployment. This observation corresponds to the notion, at least in a reduced-form sense,

**Figure 1  Estimated Coefficients**
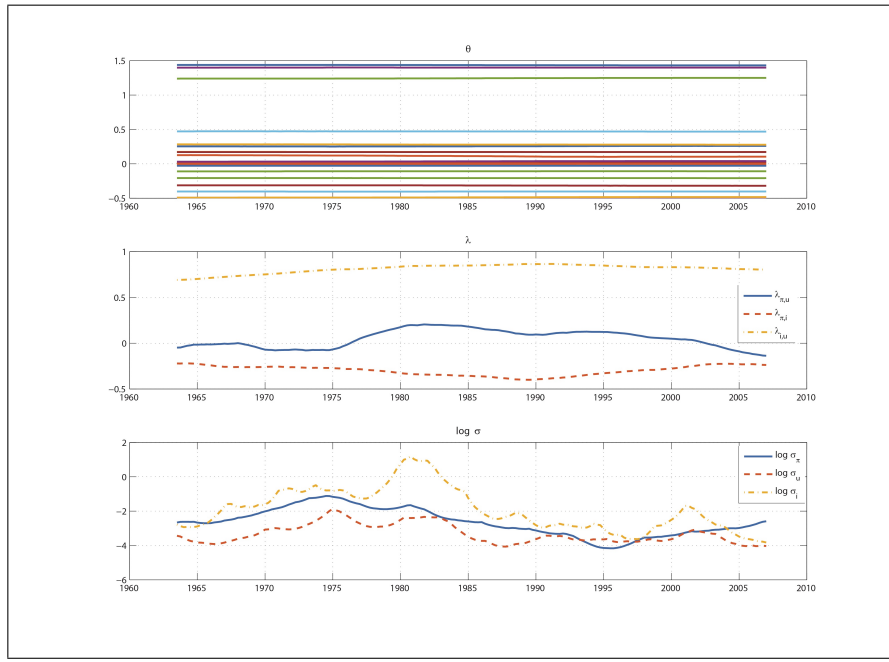


that the interest rate and unemployment move in the same direction while the interest rate and inflation rate move in the opposite direction.

The coefficient $\lambda_{\pi,u}$ for the relationship between inflation and unemployment in the middle panel exhibits more variation. It is positive from 1976 until 2002 and negative before and after. Despite uncertainty surrounding this estimate (not reported), it reveals changes in how unemployment and inflation have interacted over the sample period. This observation is of particular interest since the relationship between these two variables is sometimes described as the Phillips curve, which may embody a trade-off for the conduct of monetary policy. That this trade-off apparently changed in the late 1970s and again in the early 2000s is noteworthy. Finally, the lower panel of Figure 1 depicts the series for the elements of the $\Sigma_t$, which is a diagonal matrix. Movements in these terms indicate the extent to which volatility of the estimated errors has changed. The most variation is attributed to the interest rate, followed by the inflation rate.

Figure 1 summarizes all coefficient estimates $\theta_t$ from the TVP-VAR with stochastic volatility in a comprehensive manner. The lesson to take away from this is that almost all of the time variation in the

post-World War II history of the three variables appears to be due to stochastic volatility and not to changes in the lag coefficients. This observation is thus conceptually in line with the argument presented in Sims and Zha (2006), who use a Markov-switching VAR and also attribute changes in the behavior of the U.S. business cycle to regime changes in the shocks.

However, we want to raise some caveats for this interpretation. First, the relative importance of variations in the shocks versus changes in the parameters is a long-standing issue in econometrics, ranging from test for structural change (Lubik and Surico 2010) to the proper conditioning of state-space models including unobserved components (Stock and Watson 2003). Disentangling the relative importance of time-variation in the shocks and in lag coefficients is a challenge that a Bayesian approach has not overcome, but the judicious use of priors gives some structure to the issue. Specifically, the choice of an initial prior is informed by a pre-sample analysis, whereby the data stem from the same underlying data-generating process as the latter part of the sample.

Second, there is a concern that TVP-VARs with SV have a tendency to attribute time variation in the data to the stochastic volatility part of the model and not to the lag coefficients. In a simple example above, we argue that the inclusion of stochastic volatility is necessary to avoid a pitfall in the opposite direction. Lubik, Matthes, and Owens (2016) address this aspect in a simulation study based on an underlying nonlinear model and judge that a TVP-VAR does in fact come to the right conclusion as to the sources of time variation, but that a judicious choice of prior is crucial.

The second set of results are reported in Figure 2. These are the impulse response functions of inflation, unemployment, and the interest rate itself to a unit, that is, a 1 percentage point, increase in the three-month nominal rate bond rate. As discussed above, there are impulse responses functions at every single data point, so reporting the full set becomes a challenge. We therefore pick three dates from each decade that are associated with, respectively, the height of a deep recession, the onset of the Volcker disinflation, and at the early stages of a long expansion: 1975:Q1, 1981:Q3, and 1996:Q1. For identification purposes, the variables are in the order: inflation, unemployment, and the interest rate. This implies that the interest rate has no contemporaneous effect on inflation and unemployment, but that it responds contemporaneously to shocks in these two variables. In discussing the result, we focus on the effects of monetary policy shocks.

Figure 2 shows that the impulse responses are remarkably similar across all three time periods. This has already been indicated by the

**Figure 2 Impulse Response Functions**
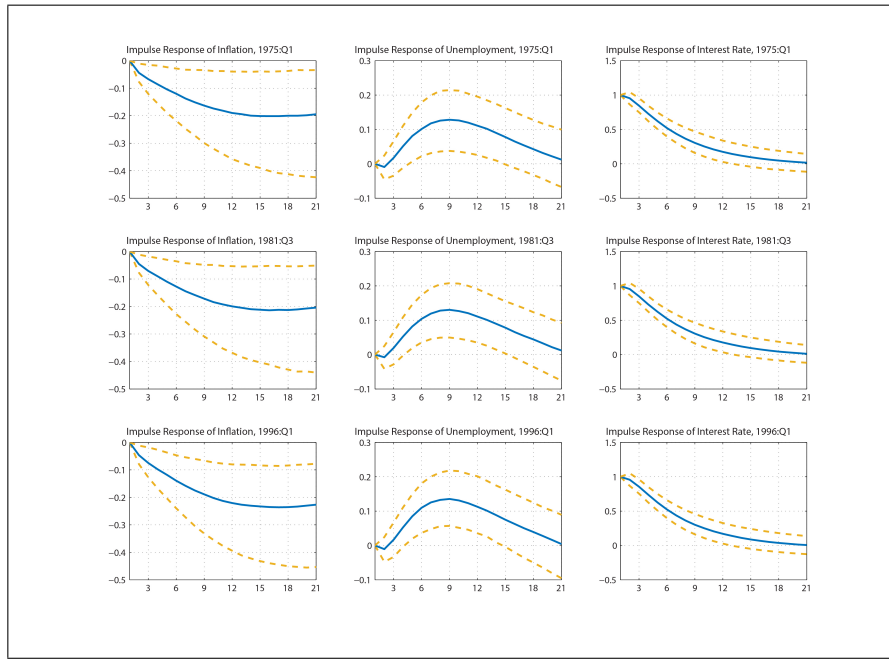


observation from Figure 1 that the estimated lag coefficients exhibit virtually no time variation. Since the impulse responses are functions of the lag coefficients, this clearly carries over. The structural responses are also functions of the matrix $H_t$ and therefore related to the factors of the reduced-form error covariance matrix, $\Lambda_t^{-1}$ and $\Sigma_t$, which show more variation; yet, this does not carry over to the impulse responses despite the sign change of the elements of $\Lambda_t^{-1}$.

Following a unit innovation, the interest rate returns slowly over time to its long-run level, which it reaches after five years. The response is fairly tightly estimated based on the 90 percent coverage regions. The interest rate's own response in the last column of the figure is very much the same in all periods. On impact, the response of the unemployment rate to a contractionary interest rate shock is zero by construction. Afterward, unemployment starts to rise slowly until hitting a peak around the two-year mark. It returns to its starting value after five years. The unemployment response is much less precisely estimated, with zero included in the coverage region for the first year after impact. Again, the responses across episodes are remarkably similar. An additional point to note is that the median extent of a 1

percentage point interest rate rise is a 0.12 percentage point increase in the unemployment rate. Finally, the interest rate hike reduces inflation over time with a fairly wide coverage region and very similar responses in each of the three time periods.

## 4.   CONCLUSION

This article discusses and reviews the concept and the methodology of time-varying parameter VARs. This class of empirical models has proved to be a flexible and comprehensive approach to capturing the dynamics of macroeconomic series. We focus on the specification and implementation of TVP-VARs in a Bayesian framework since it offers unique computational challenges. To this effect, we present the Gibbs sampler as a convenient and adaptable method for inference. We illustrate the approach by means of a simple example that estimates a small-scale TVP-VAR for the United States.

The TVP-VAR literature is still in its infancy, and there are several issues we plan to address in further detail in a companion article to the present one. Identification of structural shocks is a key element of time-series analysis. The application in the present article uses a simple, yet widely used, recursive identification scheme that is not without its problems. Alternative identification schemes, such as long-run restrictions and sign restrictions, warrant additional consideration although they present unique challenges in a TVP-VAR with SV context. A second issue is to what extent TVP-VARs are able to capture a wide variety of nonlinear behavior in macroeconomic time series, especially when compared to alternative methods, such as regime-switching VARs.

## REFERENCES

Amir-Ahmadi, Pooyan, Christian Matthes, and Mu-Chun Wang. 2016. "Drifts and Volatilities under Measurement Error: Assessing Monetary Policy Shocks over the Last Century." *Quantitative Economics,* forthcoming.

Benati, Luca, and Thomas A. Lubik. 2014. "Sales, Inventories, and Real Interest Rates: A Century of Stylized Facts." *Journal of Applied Econometrics* 29 (November/December): 1210–22.

Benati, Luca, and Paolo Surico. 2009. "VAR Analysis and the Great Moderation." *American Economic Review* 99 (September): 1636–52.

Canova, Fabio, and Fernando J. Perez-Forero. 2015. "Estimating Overidentified, Nonrecursive, Time-Varying Coefficients Structural Vector Autoregressions." *Quantitative Economics* 6 (July): 359–84.

Canova, Fabio, Filippo Ferroni, and Christian Matthes. 2015. "Approximating Time Varying Structural Models with Time Invariant Structures." Federal Reserve Bank of Richmond Working Paper 15-10 (September).

Canova, Fabio, and Luca Gambetti. 2009. "Structural Changes in the US Economy: Is There a Role for Monetary Policy?" *Journal of Economic Dynamics and Control* 33 (February): 477–90.

Carter, C. K., and R. Kohn. 1994. "On Gibbs Sampling for State Space Models." *Biometrika* 81 (September): 541–53.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary Policy Shocks: What Have We Learned and To What End?" In *Handbook of Macroeconomics, vol. 1*, edited by John B. Taylor and Michael Woodford. North Holland: Elsevier, 65–148.

Cogley, Timothy, and Thomas J. Sargent. 2002. "Evolving Post-World War II U.S. Inflation Dynamics." In *NBER Macroeconomics Annual 2001, vol. 16*, edited by Ben S. Bernanke and Kenneth Rogoff. Cambridge, Mass.: MIT Press, 331–88.

Cogley, Timothy, and Thomas J. Sargent. 2005. "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S." *Review of Economic Dynamics* 8 (April): 262–302.

Cogley, Timothy, Argia Sbordone, and Christian Matthes. 2015. "Optimized Taylor Rules for Disinflation When Agents are Learning." *Journal of Monetary Economics* 72 (May): 131–47.

Del Negro, Marco, and Giorgio Primiceri. 2015. "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum." *Review of Economic Studies*, forthcoming.

Doh, Taeyoung, and Michael Connolly. 2012. "The State Space Representation and Estimation of a Time-Varying Parameter VAR with Stochastic Volatility." Federal Reserve Bank of Kansas City Working Paper 12-04 (July).

Engle, Robert F. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (July), 987–1008.

Gelman, Andrew, et al. 2014. *Bayesian Data Analysis.* Third Edition. Boca Raton: CRC Press.

Fernandez-Villaverde, Jesus, et al. 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97 (June): 1021–26.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib. 1998. "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies* 65 (July): 361–93.

Koop, Gary, and Demetrios Korobilis. 2010. "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics." Manuscript.

Lubik, Thomas A., Christian Matthes, and Andrew Owens. 2016. "Beveridge Curve Shifts and Time-Varying Parameter VARs." Manuscript.

Lubik, Thomas A., and Frank Schorfheide. 2004. "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* 94 (March): 190–217.

Lubik, Thomas A., and Paolo Surico. 2010. "The Lucas Critique and the Stability of Empirical Models." *Journal of Applied Econometrics* 25 (January/February): 177–94.

Nakajima, Jouchi. 2011. "Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications." IMES Bank of Japan Discussion Paper 2011-E-9 (March).

Primiceri, Giorgio E. 2005. "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies* 72 (July): 821–52.

Robert, Christian, and George Casella. 2004. *Monte Carlo Statistical Methods.* Second Edition. New York: Springer Verlag.

Rondina, Francesca. 2013. "Time Varying SVARs, Parameter Histories, and the Changing Impact of Oil Prices on the US Economy." Manuscript.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (January): 1–48.

Sims, Christopher A. 2002. "Comment on Cogley and Sargent's 'Evolving Post-World War II U.S. Inflation Dynamics.'" In *NBER Macroeconomics Annual 2001, vol. 16,* edited by Ben S. Bernanke and Kenneth Rogoff. Cambridge, Mass.: MIT Press, 373–79.

Sims, Christopher A., and Tao Zha. 2006. "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96 (March): 54–81.

Stock, James H., and Mark M. Watson. 2003. "Has the Business Cycle Changed and Why?" In *NBER Macroeconomics Annual 2002, vol. 17,* edited by Mark Gertler and Kenneth Rogoff. Cambridge, Mass.: MIT Press, 159–230.

Uhlig, Harald. 1997. "Bayesian Vector Autoregressions with Stochastic Volatility." *Econometrica* 65 (January): 59–74.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics.* New York: J. Wiley and Sons, Inc.