working paper

18 11

A Class of Time-Varying Parameter Structural VARs for Inference under Exact or Set Identification

Mark Bognanni

FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working Paper 18-11 September 2018

A Class of Time-Varying Parameter Structural VARs for Inference under Exact or Set Identification

Mark Bognanni

This paper develops a new class of structural vector autoregressions (SVARs) with time-varying parameters, which I call a drifting SVAR (DSVAR). The DSVAR is the first structural time-varying parameter model to allow for internally consistent probabilistic inference under exact–or set–identification, nesting the widely used SVAR framework as a special case. I prove that the DSVAR implies a reduced-form representation, from which structural inference can proceed similarly to the widely used two-step approach for SVARs: beginning with estimation of a reduced form and then choosing among observationally equivalent candidate structural parameters via the imposition of identifying restrictions. In a special case, the implied reduced form is a tractable known model for which I provide the first algorithm for Bayesian estimation of all free parameters. I demonstrate the framework in the context of Baumeister and Peersman's (2013b) work on time variation in the elasticity of oil demand.

Keywords: structural vector autoregressions, time-varying parameters, Gibbs sampling, stochastic volatility, Bayesian inference.

JEL codes: C11, C15, C32, C52, E3, E4, E5.

Suggested citation: Bognanni, Mark. 2018. "A Class of Time-Varying Parameter Structural VARs for Inference under Exact or Set Identification." Federal Reserve Bank of Cleveland, Working Paper no. 18-11. https://doi.org/10.26509/frbc-wp-201811.

Mark Bognanni is at the Federal Reserve Bank of Cleveland (markbognanni@gmail.com; http://markbognanni.com). The author is particularly indebted to Juan Rubio-Ramírez, Harald Uhlig, and Dan Waggoner for multiple detailed conversations. He thanks Christiane Baumeister for helpful conversations (and because her code runs). He also thanks Dario Caldara, Domenico Giannone, Yuriy Gorodnichenko, Jim Hamilton, Eric Leeper, Christian Matthes, Karel Mertens, Frank Schorfheide, and Mike West, as well as his Cleveland Fed colleagues Dionissi Aliprantis, Todd Clark, Ed Knotek, Ellis Tallman, and Saeed Zaman, for helpful comments and conversations. He thanks John Zito for outstanding research assistance.

1. Introduction

The last two decades of research on structural vector autoregressions (SVARs) have developed largely in pursuit of relaxing two constraints: constant model parameters and dogmatic identifying restrictions. Research focused on relaxing the assumption of constant parameters has followed from the time-varying parameter VAR with stochastic volatility (VAR-TVP-SV) of Cogley and Sargent (2005) and Primiceri (2005) and the Markov-switching (MS-VAR) model developed in Sims and Zha (2006) and Sims, Waggoner, and Zha (2008).[1] Research focused on achieving structural inference while relaxing traditional "zero restrictions" has followed from the seminal contributions of Faust (1998), Canova and De Nicolo (2002), and Uhlig (2005). The most widely used variation on these alternative identifying approaches yields set identification for the objects of interest by imposing restrictions on the sign of certain impulse responses, and hence they are often referred to as "sign restrictions." The recent growth in popularity of both types of extensions would be difficult to overstate. To date, however, the research agendas on time-varying parameters and set identification have lived largely separate lives.

This paper's contribution is to be the first to develop a class of structural time-varying-parameter vector autoregressions from which researchers can obtain internally consistent probabilistic inference under exact–or set–identification. The structural TVP model I propose uses laws of motion for the time-varying parameters that fundamentally differ from those used in the literature previously. The new model I provide for the evolution of the structural parameters yields *sequences* of time-varying structural parameters that remain observationally equivalent under orthogonal rotations, analogous to the standard SVAR with constant parameters. Speaking somewhat loosely, approaches to identifying SVARs then carry over in a natural way to the time-varying parameter model. In short, one might alternatively summarize this paper's key contribution as providing an SVAR with time-varying parameters that is amenable to the wholesale extension of the widely used methods developed in Rubio-Ramírez, Waggoner, and Zha

[1] See Cogley and Sargent (2001) for an earlier version of the VAR-TVP model without stochastic volatility. The desirability of model extensions in this direction, and at least a partial description of how one might formulate such models, goes back to Doan, Litterman, and Sims (1984) and Sims (1993).

(2010).

A popular approach to inference in SVARs, resulting from the identification problem, is to first estimate a reduced form implied by the structural model. Structural inference then proceeds by imposing certain identifying assumptions as a sort of post-processing stage in which the researcher chooses among the observationally equivalent candidate truths hiding inside the reduced form. This is the first paper in the literature to write down a time-varying structural system and provide conditions under which it implies a reduced-form. Then the path to structural inference can proceed analogously to that with which researchers are already familiar for SVARs. In a notable special case, the reducedform implied by the structural model is highly tractable and relatively well-developed in the Bayesian statistics literature. I also provide what I believe to be the first algorithm for the fully Bayesian estimation of *all* of the model's free parameters. Furthermore, the MCMC algorithm is fast enough, even in high-level programming languages such as `MATLAB`, for the estimation of at least medium-sized VARs to be practical.

At this stage, one might wonder why a new model is needed at all. Could one not simply use a piecemeal combination of the existing VAR-TVP-SV models and the set identification methods developed for constant-parameter VARs? Indeed, in recent years researchers have explored such an approach based on first estimating the model of Primiceri (2005) followed by the use of Uhlig-like methods, on a period-by-period basis, on the estimated time-varying coefficients from the first stage.[2] However, such an approach is subject to two key shortcomings in the interpretation of its results. First, no model for the time-varying structural parameters is ever proposed that would rationalize the various orthogonally rotated parameters as observationally equivalent candidate truths. Second, inference in the model of Primiceri (2005) depends on the ordering of variables in the VAR. Hence, an n-variable dynamic system admits $n!$ distinct estimates of the time-varying parameters in the first stage, which form the key input into the set identification procedure.[3] The dependence on variable ordering then passes

[2]See Canova and Gambetti (2009), Hofmann, Peersman, and Straub (2012), Baumeister and Peersman (2013b), Baumeister and Peersman (2013a), and Baumeister and Benati (2013) for examples.

[3]Indeed, the potential sensitivity of results to variable ordering is known and acknowledged in both Cogley and Sargent (2005) and Primiceri (2005). See also Section 8 of Fox and West

through all the way to structural inference when applying the set identification algorithm to the estimated parameters. In Section 4 I show that in practical applications this property can easily become much more than just an intellectual curiosity.

This paper builds off of a number of papers more familiar to Bayesian statisticians than to economists. In a special case of the structural model, the implied reduced form is known to Bayesian statisticians as a dynamic linear model (DLM) with discounted Wishart stochastic volatility (DLM-DWSV). Variants of the DLM with a constant covariance matrix have been used to model financial time-series since at least Quintana and West (1987), while the discounted Wishart stochastic volatility process was formalized as a valid probability model by Uhlig (1994) and Uhlig (1997). Prado and West (2010) give the most thorough treatment to date of the complete model.[4] Koop and Korobilis (2013) consider forecasting with a model similar to the reduced-form DLM-DWSV but without fully Bayesian likelihood-based estimation of the model parameters.

From here the rest of the paper proceeds as follows. In Section 2 I review the current frameworks for exact and set identification in constant-parameter SVARs to emphasize the parallelism with my time-varying parameter extension. In Section 3 I describe the key issues involved in developing a framework with time-varying parameters that can be analyzed in a fashion analogous to constant-parameter models. In Section 4 I motivate this paper's developments by demonstrating the shortcomings of the current approach to set identification in time-varying parameter systems. In Section 5 I present my structural VAR with time-varying parameters and the reduced-form model it implies. In Section 6 I describe the MCMC algorithm for estimating the reduced-form model. In Section 7 I confront the model and estimation procedure with an empirical application from the literature, that of Baumeister and Peersman (2013b) on the time-varying effects of oil supply shocks. In Section 8 I conclude.

Notation. Before moving on, I introduce a few notational conventions used throughout the paper. Scalars are styled as lowercase letters in normal weight, as

(2014) for a discussion of this issue. Note that both Cogley and Sargent (2005) and Primiceri (2005) work with small dynamic systems of three variables and two lags, which make it feasible to check robustness against all six possible orderings.

[4]See also Harrison and West (1997) and Quintana, Lourdes, Aguilar, and Liu (2003).

in β. Vectors are styled as lowercase letters in bold, as in $\mathbf{y}$. Matrices are styled as uppercase letters in bold, as in $\mathbf{A}$. The value of a time-varying object at a particular time t is referenced by including a subscript, as in $\mathbf{A}_t$ and the collection of all values between the two points in time t and $t+k$ is denoted with a ":" in the subscript, as in $\mathbf{A}_{t:t+k} \equiv \{\mathbf{A}_t, \mathbf{A}_{t+1}, \dots, \mathbf{A}_{t+k-1}, \mathbf{A}_{t+k}\}$.

With respect to some special matrices, the $n \times n$ identity matrix is denoted $\mathbf{I}_n$ and $\mathbf{0}_{m,n}$ denotes an $m \times n$ matrix of zeroes. $\mathcal{O}_n$ denotes the group of $n \times n$ orthogonal matrices. I often refer to symmetric positive definite matrices as "SPD." Letting $\mathbf{\Sigma}$ be SPD, $\underline{h}(\mathbf{\Sigma})$ denotes the unique lower triangular matrix with positive diagonal elements for which $\underline{h}(\mathbf{\Sigma})\underline{h}(\mathbf{\Sigma})' = \mathbf{\Sigma}$, while the notation $\mathbf{\Sigma}^{1/2}$ denotes the unique SPD matrix such that $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$.

Well-known probability distributions are denoted with uppercase letters in normal weight, as in $N(0,1)$ to denote the standard normal distribution. The notation $N(\mathbf{m}, \mathbf{\Sigma})$ refers to the multivariate normal distribution, while $N(\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ refers to a matrix-variate normal distribution with $m \times n$ mean matrix $\mathbf{M}$, $m \times m$ row covariance matrix $\mathbf{\Sigma}_r$, and $n \times n$ column covariance matrix $\mathbf{\Sigma}_c$.

2. A Crash Course in SVARs

To fix ideas, I start by describing the most widely used framework for inference in constant-parameter SVARs. My description draws heavily from Rubio-Ramírez et al. (2010), Del Negro and Schorfheide (2011), and Arías, Rubio-Ramírez, and Waggoner (2018), to which I refer readers for further details.

2.1 SVAR: The Structural Model

The structural model assumes that n observable, endogenous economic variables evolve according to a linear relationship between each variable and the contemporaneous and lagged values of all variables and a constant term. In other words, I assume that the $(n \times 1)$ vector of observables at time t, denoted $\mathbf{y}_t$, is realized according to a structural vector autoregression written as

$$(1) \qquad \mathbf{y}_t'\mathbf{A} = \sum_{\ell=1}^{p} \mathbf{y}_{t-\ell}'\mathbf{F}_{(\ell)} + \mathbf{c} + \boldsymbol{\varepsilon}_t', \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n,1}, \mathbf{I}_n), \quad \text{for } 1 \leq t \leq T,$$

where $\boldsymbol{\varepsilon}_t$ is an $(n \times 1)$ vector of exogenous and mutually orthogonal structural shocks. The integer p is the number of lags of observables pertinent to the structural representation of the dynamic system. The matrices $\mathbf{A}$ and $\mathbf{F}_{(\ell)}$ for $0 \leq \ell \leq p$ are each $n \times n$ and $\mathbf{c}$ is a $(1 \times n)$ vector. I also assume that $\mathbf{A}$ is invertible.

To make the subsequent exposition more concise, I define $m \equiv p{\cdot}n{+}1$, the $(m\times n)$ matrix $\mathbf{F} \equiv [\mathbf{F}'_{(1)}, \dots, \mathbf{F}'_{(p)}, \mathbf{c}']'$, and the $(m \times 1)$ vector $\mathbf{x}_t \equiv [\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, 1]'$. One can then write the model in equation (1) compactly as

$$\mathbf{y}'_t\mathbf{A} = \mathbf{x}'_t\mathbf{F} + \boldsymbol{\varepsilon}'_t\,, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n,1}, \mathbf{I}_n)\,, \quad \text{for } 1 \leq t \leq T\,. \tag{2}$$

I refer to $(\mathbf{A}, \mathbf{F})$ as the structural parameters because they determine the evolution of the endogenous economic variables in response to the mutually orthogonal exogenous disturbances $\boldsymbol{\varepsilon}_t$. The objects of interest to the economist are either particular elements of $(\mathbf{A}, \mathbf{F})$ or functions thereof, such as impulse responses or variance decompositions.

From the normality of $\boldsymbol{\varepsilon}_t$, the density for the vector $\mathbf{y}'_t\mathbf{A}$ has the distribution $N(\mathbf{y}'_t\mathbf{A}|\mathbf{x}'_t\mathbf{F}, \mathbf{I}_n)$. When $\mathbf{A}^{-1}$ exists, the density of $\mathbf{y}'_t$ can be found by transforming the random vector $\mathbf{y}'_t\mathbf{A}$ via the right-multiplication by $\mathbf{A}^{-1}$, and finding the distribution of $p\big((\mathbf{y}'_t\mathbf{A})\mathbf{A}^{-1}|\mathbf{A}, \mathbf{F}, \mathbf{x}_t\big) = p(\mathbf{y}'_t|\mathbf{A}, \mathbf{F}, \mathbf{x}_t)$. From well-known properties of the multivariate normal distribution under affine transformations,

$$p(\mathbf{y}'_t|\mathbf{A}, \mathbf{F}, \mathbf{x}_t) = N\big(\mathbf{y}'_t \,|\, \mathbf{x}'_t\mathbf{F}\mathbf{A}^{-1},\, (\mathbf{A}\mathbf{A}')^{-1}\big)\,. \tag{3}$$

and the likelihood is given by $p(\mathbf{y}_{1:T}|\mathbf{A}, \mathbf{F}) = \prod_{t=1}^{T} p(\mathbf{y}'_t|\mathbf{A}, \mathbf{F}, \mathbf{x}_t)$.

In the absence of further restrictions, the space of possible values for $(\mathbf{A}, \mathbf{F})$ is the subset of $\mathbb{R}^{mn+n^2}$ for which $\mathbf{A}$ is invertible, and I refer to this space as $\mathbb{S}$. The structural model then consists of a delineation of the unobservables $(\mathbf{A}, \mathbf{F})$, the space of admissible values for those unobservables $\mathbb{S}$, and the data density for $\mathbf{y}_{1:T}$ conditional on the unobservables.

2.2 The Identification Problem in SVARs

Following Rothenberg (1971), I consider two parameter points of a model to be observationally equivalent if and only if they imply the same distribution of $\mathbf{y}_{1:T}$.

Definition 1 (Observational equivalence)**.** *The points* $(\mathbf{A}, \mathbf{F})$ *and* $(\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}})$ *are observationally equivalent if and only if* $p\big(\mathbf{y}_{1:T}|\mathbf{A}, \mathbf{F}\big) = p\big(\mathbf{y}_{1:T}|\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}}\big)$.

Since no data can ever distinguish between the two points, I say that a model is not identified at $(\mathbf{A}, \mathbf{F})$ if an observationally equivalent $(\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}})$ exists.

For the structural model, it is well-known in the literature that, given any point $(\mathbf{A}, \mathbf{F}) \in \mathbb{S}$, the alternative point $(\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}}) \in \mathbb{S}$ is observationally equivalent if (and only if) there exists a $\mathbf{Q} \in \mathcal{O}_n$ such that $(\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}}) = (\mathbf{AQ}, \mathbf{FQ})$.[5] Since, for any $(\mathbf{A}, \mathbf{F})$, there are as many such points as there are matrices in $\mathcal{O}_n$, it is apparent that the parameters $(\mathbf{A}, \mathbf{F})$ are not identified.

2.3 A Useful Reparameterization of SVARs

Although $(\mathbf{A}, \mathbf{F})$ are not identified, one can identify certain *combinations* of parameters in $(\mathbf{A}, \mathbf{F})$. This fact becomes particularly apparent under a reparameterization of the structural model. Arías et al. (2018) show that the structural model can be reparameterized as $(\mathbf{B}, \mathbf{H}, \mathbf{Q}) = f(\mathbf{A}, \mathbf{F})$ where

$$f(\mathbf{A}, \mathbf{F}) = \Big(\underbrace{\mathbf{FA}^{-1}}_{\mathbf{B}}, \underbrace{\mathbf{AA}'}_{\mathbf{H}}, \underbrace{\underline{h}(\mathbf{AA}')^{-1}\mathbf{A}}_{\mathbf{Q}} \Big). \tag{4}$$

$\mathbf{H}$ is symmetric positive definite and $\mathbf{Q} \in \mathcal{O}_n$.[6][7] Letting $\mathbb{H}(n)$ denote the space of $n \times n$ symmetric positive definite matrices, the space of possible values for $(\mathbf{B}, \mathbf{H})$ is $\mathbb{R}^{m \times n} \times \mathbb{H}(n)$ and I refer to this space as $\mathbb{D}$. The space of possible values for $(\mathbf{B}, \mathbf{H}, \mathbf{Q})$ is then $\mathbb{D} \times \mathcal{O}_n$ and I refer to this space as $\mathbb{DO}$. The mapping f is invertible, with f^{-1} given by

$$f^{-1}(\mathbf{B}, \mathbf{H}, \mathbf{Q}) = \Big(\underbrace{\underline{h}(\mathbf{H})\,\mathbf{Q}}_{\mathbf{A}}, \underbrace{\mathbf{B}\,\underline{h}(\mathbf{H})\,\mathbf{Q}}_{\mathbf{F}} \Big). \tag{5}$$

[5] See, for example, Rubio-Ramírez et al. (2010).

[6] Also see Del Negro and Schorfheide (2011).

[7] Calling f a "reparameterization" may appear curious at first since the $n \times n$ and $m \times n$ matrices of $(\mathbf{A}, \mathbf{F})$ are transformed into $n \times n$ and $m \times n$ matrices $(\mathbf{B}, \mathbf{H})$ plus an additional $n \times n$ matrix $\mathbf{Q}$. The ostensible inconsistency is resolved by noting that $\mathbf{H}$ is symmetric, and hence has only $n(n+1)/2$ functionally independent elements, and $\mathbf{Q} \in \mathcal{O}_n$, and hence has only $n(n-1)/2$ functionally independent elements. Thus, the total number of functionally independent elements is the same under either parameterization.

The reparameterization under f is attractive for a number of reasons. First, it makes the identification problem transparent. Under the substitutions of f^{-1} one can write the model as

$$\text{(6)} \qquad \mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \varepsilon_t'\mathbf{Q}'\underline{h}(\mathbf{H})^{-1'}, \quad \varepsilon_t \sim N(\mathbf{0}_{n,1}, \mathbf{I}_n), \quad \text{for } 1 \le t \le T,$$

with conditional likelihood from equation (3) becoming

$$\text{(7)} \quad p(\mathbf{y}_t'|\mathbf{B}, \mathbf{H}, \mathbf{Q}, \mathbf{x}_t) = N\left(\mathbf{y}_t'|\mathbf{x}_t'\mathbf{B}, \left(\underline{h}(\mathbf{H})\mathbf{Q}\mathbf{Q}'\underline{h}(\mathbf{H})'\right)^{-1}\right) = N(\mathbf{y}_t'|\mathbf{x}_t'\mathbf{B}, \mathbf{H}^{-1}),$$

and likelihood given by

$$\text{(8)} \qquad p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H}, \mathbf{Q}) = \prod_{t=1}^{T} \overbrace{p(\mathbf{y}_t'|\mathbf{B}, \mathbf{H}, \mathbf{Q}, \mathbf{x}_t)}^{\substack{= N(\mathbf{y}_t'|\mathbf{x}_t'\mathbf{B}, \mathbf{H}^{-1}) \\ \text{from eqn (7)}}} = p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H}).$$

Since the likelihood function can be written entirely without recourse to $\mathbf{Q}$ it is apparent that $\mathbf{Q}$ is not identifiable.

Priors and Posteriors. A second attractive feature of the parameterization under f is that it allows for a prior cleanly separated into the researcher's beliefs for objects that the data can identify, $(\mathbf{B}, \mathbf{H})$, and the researcher's beliefs for objects that the data cannot identify, $\mathbf{Q}$: a prior $p(\mathbf{B}, \mathbf{H}, \mathbf{Q})$ can in general be written as

$$\text{(9)} \qquad p(\mathbf{B}, \mathbf{H}, \mathbf{Q}) = p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{B}, \mathbf{H}),$$

where the density $p(\mathbf{Q}|\mathbf{B}, \mathbf{H})$ articulates what the researcher will infer about the structural parameters conditional on the values of the objects the data can inform.[8]

To further clarify how $\mathbf{Q}$ is handled under Bayesian inference, observe that

[8]The functional form of the likelihood is such that $p(\mathbf{B}, \mathbf{H})$ can be conjugate, and also *natural* conjugate. Natural conjugate priors have the interpretation of the shape of the likelihood function under some notional data. Priors of this class provide a useful device for disciplining prior formulation by requiring the researcher's prior beliefs to be consistent with the shape of the likelihood under some conceivable data.

the posterior for the unobservables $(\mathbf{B}, \mathbf{H}, \mathbf{Q})$ is given by

$$(10) \qquad p(\mathbf{B}, \mathbf{H}, \mathbf{Q}|\mathbf{y}_{1:T}) = p(\mathbf{y}_{1:T})^{-1} \cdot p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{B}, \mathbf{H}) \cdot \overbrace{p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H}, \mathbf{Q})}^{p(\mathbf{y}_{1:T}|\mathbf{B},\mathbf{H})}$$

where

$$(11) \qquad p(\mathbf{y}_{1:T}) = \int_{(\mathbf{B},\mathbf{H},\mathbf{Q})} p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H}, \mathbf{Q}) \, d\mathbf{B} \, d\mathbf{H} \, d\mathbf{Q}$$

The term $p(\mathbf{y}_{1:T})$, known as the marginal data density or marginal likelihood, is an important Bayesian notion of model fit. Since the likelihood does not depend on $\mathbf{Q}$ one can simplify the expression for $p(\mathbf{y}_{1:T})$ as

$$(12) \qquad p(\mathbf{y}_{1:T}) = \int_{(\mathbf{B},\mathbf{H})} \underbrace{\left(\int_{\mathbf{Q}} p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \, d\mathbf{Q} \right)}_{=1} p(\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H}) \, d\mathbf{B} \, d\mathbf{H}$$

Hence, the marginal data density of the model does not depend on the particular choice of $p(\mathbf{Q}|\mathbf{B}, \mathbf{H})$. The posterior distribution can then be written as

$$(13) \qquad p(\mathbf{B}, \mathbf{H}, \mathbf{Q}|\mathbf{y}_{1:T}) = p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot \frac{p(\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H})}{\int_{(\mathbf{B},\mathbf{H})} p(\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H}) \, d(\mathbf{B}, \mathbf{H})}.$$

Noting that the second term is precisely the marginal posterior of $(\mathbf{B}, \mathbf{H})$, the posterior can be written as

$$(14) \qquad p(\mathbf{B}, \mathbf{H}, \mathbf{Q}|\mathbf{y}_{1:T}) = p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{B}, \mathbf{H}|\mathbf{y}_{1:T}).$$

Equation (14) makes clear the way Bayesian inference will incorporate information from the data to make inference about the structural parameters: the data will update beliefs about the elements of $(\mathbf{B}, \mathbf{H})$ directly, while beliefs about $\mathbf{Q}$ are updated by the data only indirectly via the information the data provide about $(\mathbf{B}, \mathbf{H})$. The conditional posterior of $\mathbf{Q}$ is simply the conditional prior.

2.4 Identifying Restrictions in SVARs

Considering inference about the entirety of the space of $(\mathbf{B}, \mathbf{H}, \mathbf{Q})$ will typically preclude the possibility of making any economically substantive statements. To narrow the scope of economic conclusions from the data, researchers typically impose additional restrictions on the space of potential structural parameters. In general, one might represent such restrictions as a set $\mathcal{R} \subseteq \mathbb{D}\mathbb{O}$ and thus statistical inference must respect the requirement that $(\mathbf{B}, \mathbf{H}, \mathbf{Q}) \in \mathcal{R}$.

From the standpoint of Bayesian inference, one can mechanically construct a prior that respects $\mathcal{R}$ as

$$p^{\mathcal{R}}(\mathbf{B},\mathbf{H}, \mathbf{Q}) \propto p(\mathbf{B}, \mathbf{H}, \mathbf{Q}) \cdot I\left\{(\mathbf{B}, \mathbf{H}, \mathbf{Q}) \in \mathcal{R}\right\} \tag{15}$$

which can in general be factored into a marginal and conditional similarly to the unrestricted case, as

$$\begin{aligned} p^{\mathcal{R}}(\mathbf{B}, \mathbf{H}, \mathbf{Q}) \propto\ & p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot I\left\{\mathbf{Q} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B}, \mathbf{H})\right\} \\ & \cdot p(\mathbf{B}, \mathbf{H}) \cdot I\left\{(\mathbf{B}, \mathbf{H}) \in \mathcal{R}_{\mathbf{B},\mathbf{H}}\right\} \end{aligned} \tag{16}$$

where

$$\mathcal{R}_{\mathbf{B},\mathbf{H}} \equiv \left\{(\mathbf{B}, \mathbf{H}) : \exists\, \mathbf{Q} \text{ for which } (\mathbf{B}, \mathbf{H}, \mathbf{Q}) \in \mathcal{R}\right\} \tag{17}$$

$$\mathcal{R}_{\mathbf{Q}}(\mathbf{B}, \mathbf{H}) \equiv \left\{\mathbf{Q} : (\mathbf{B}, \mathbf{H}, \mathbf{Q}) \in \mathcal{R}\right\}. \tag{18}$$

The two most widely used approaches to identification in the SVAR literature are exact and set identification. I conceive of both types of restrictions as "exact-or-less" in the sense of implying unrestricted support for $(\mathbf{B}, \mathbf{H})$ on $\mathbb{D}$, and hence $I\{(\mathbf{B}, \mathbf{H}) \in \mathcal{R}_{\mathbf{B},\mathbf{H}}\} = 1$ for almost all $(\mathbf{B}, \mathbf{H})$. Defining the restricted prior for $\mathbf{Q}$ as

$$p^{\mathcal{R}}(\mathbf{Q}|\mathbf{B}, \mathbf{H}) = \frac{p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot I\{\mathbf{Q} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B}, \mathbf{H})\}}{\int_{\mathbf{Q}} p(\mathbf{Q}|\mathbf{B}, \mathbf{H}) \cdot I\{\mathbf{Q} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B}, \mathbf{H})\} d\mathbf{Q}}, \tag{19}$$

and substituting that expression into equation (14), the posterior of the restricted model can be written as

$$p^{\mathcal{R}}\big(\mathbf{B},\mathbf{H},\mathbf{Q}|\mathbf{y}_{1:T}\big) = p^{\mathcal{R}}(\mathbf{Q}|\mathbf{B},\mathbf{H}) \cdot p\big(\mathbf{B},\mathbf{H}|\mathbf{y}_{1:T}\big)\,. \tag{20}$$

I call the identifying restrictions exact if $\mathcal{R}$ is such that there is a function mapping each $(\mathbf{B},\mathbf{H})$ to a single $\mathbf{Q}$. In other words, under exact identification $p^{\mathcal{R}}(\mathbf{Q}|\mathbf{B},\mathbf{H})$ is a pointmass. Intuitively, exact identification amounts to choosing a single element from the set of observationally equivalent structural parameters associated with a given $(\mathbf{B},\mathbf{H})$.[9]

Set identification, which most often takes the form of restrictions on the "signs" of structural impulse responses, amounts to defining $\mathcal{R}$ in such a way as to allow a positive measure of $\mathbf{Q}$ matrices in $\mathcal{R}_{\mathbf{Q}}(\mathbf{B},\mathbf{H})$ for almost all $(\mathbf{B},\mathbf{H})$. When working within the Bayesian paradigm, the natural next step is to place a prior over the $\mathbf{Q}$ matrices in $\mathcal{O}_n$. The most common choice in the literature is

$$p^{\mathcal{R}}(\mathbf{Q}|\mathbf{B},\mathbf{H}) \propto p(\mathbf{Q}) \cdot I\{\mathbf{Q} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B},\mathbf{H})\} \tag{21}$$

where $p(\mathbf{Q})$ is the uniform distribution over $\mathcal{O}_n$, which I denote $U(\mathcal{O}_n)$. Rubio-Ramírez et al. (2010) give an efficient algorithm for generating random draws from such priors based on combining the algorithm of Stewart (1980) with an accept-reject step.[10]

2.5 Reduced form of the SVAR

The irrelevance of $\mathbf{Q}$ for the likelihood, and of $p(\mathbf{Q}|\mathbf{B},\mathbf{H})$ for the Bayesian marginal likelihood, implies that one could fit the data just as well by formulating a model entirely without $\mathbf{Q}$. Indeed, one could consider the parameters $(\mathbf{B},\mathbf{H})$, their prior density $p(\mathbf{B},\mathbf{H})$, and the data density in equation (7), to be a model in and of itself.

---

[9]This notion of exact identification is sufficiently general to accommodate the conditions given in Rubio-Ramírez et al. (2010) (see their Theorem 5 and Algorithm 1) as well as the penalty function approach as in Uhlig (2005) and Mountford and Uhlig (2009).

[10]See Algorithm 2 in Rubio-Ramírez et al. (2010). My description of the approach to structural inference with partially identifying restrictions represents only the most commonly implemented approach in the literature; the validity of the resulting inference for objects of interest in a particular application presumes that the researcher is comfortable with the prior for the objects of interest induced by the Haar measure prior over $\mathcal{O}_n$. See Baumeister and Hamilton (2015) for an alternative approach to identification in constant-parameter SVARs.

Writing the model as

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \mathbf{u}_t', \quad \mathbf{u}_t' \sim N(\mathbf{0}_{n,1}, \mathbf{H}^{-1}), \quad \text{for } 1 \le t \le T \tag{22}$$

yields a data density $p(\mathbf{y}_t|\mathbf{B}, \mathbf{H}, \mathbf{x}_t) = N(\mathbf{y}_t|\mathbf{B}'\mathbf{x}_t, \mathbf{H}^{-1})$, which is identical to equation (7). With identical likelihoods and identical priors, the posterior of $(\mathbf{B}, \mathbf{H})$ in the VAR,

$$p(\mathbf{B}, \mathbf{H}|\mathbf{y}_{1:T}) = \frac{p(\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H})}{\int_{(\mathbf{B},\mathbf{H})} p(\mathbf{B}, \mathbf{H}) \cdot p(\mathbf{y}_{1:T}|\mathbf{B}, \mathbf{H})\, d(\mathbf{B}, \mathbf{H})}, \tag{23}$$

is precisely the *marginal* posterior of $(\mathbf{B}, \mathbf{H})$ in the reparameterized SVAR. Hence, I say that the reduced-form model is implied by the structural model.

2.6 From Reduced-Form Parameter Estimation to Structural Inference

Even when the objects of interest to the economist depend on the full set of structural parameters $(\mathbf{B}, \mathbf{H}, \mathbf{Q})$, the practice of inference often begins by estimating $(\mathbf{B}, \mathbf{H})$ in the reduced-form model. Such an approach is useful for two reasons. First, as noted in the previous section, the posterior of $(\mathbf{B}, \mathbf{H})$ under the reduced-form model is identical to the marginal posterior of $(\mathbf{B}, \mathbf{H})$ in the structural model. Hence, reduced-form estimation exactly represents a region of the structural model's posterior. From a practical perspective this is only useful insomuch as the posterior of the VAR is tractable, which is very much the case. Second, after conditioning on $(\mathbf{B}, \mathbf{H})$ the structural model's posterior of $\mathbf{Q}$ does not depend on the data.

Following estimation of $(\mathbf{B}, \mathbf{H})$ in the VAR, researchers proceed to structural inference by sampling from $p^{\mathcal{R}}(\mathbf{Q}|\mathbf{B}, \mathbf{H})$, which effectively becomes a form of post-processing of the reduced-form parameter draws. Thus the algorithm often used by researchers in practice is Algorithm 1.

Algorithm 1 - Structural posterior sampling

---

1. ***Reduced-form model estimation***

- *For* $i = 1, \dots, n_1$, *sample* $(\mathbf{B}^{(i)}, \mathbf{H}^{(i)}) \sim p(\mathbf{B}, \mathbf{H}|\mathbf{y}_{1:T})$
- *Store* $\left\{\mathbf{B}^{(i)}, \mathbf{H}^{(i)}\right\}_{i=1}^{n_1}$

2. ***Structural inference***

- *For each draw in* $\left\{\mathbf{B}^{(i)}, \mathbf{H}^{(i)}\right\}_{i=1}^{n_1}$, *sample* $\mathbf{Q}^{(i)} \sim p^{\mathcal{R}}(\mathbf{Q}|\mathbf{B}^{(i)}, \mathbf{H}^{(i)})$.

An appealing aspect of Algorithm 1 is that, by isolating the step of structural inference, researchers can assess the robustness of their results to alternative identification schemes without needing to re-run the reduced-form estimation.

2.7 The SVAR's Tractability (or The "Want Operator" for a TVP Extension)

Before moving to the time-varying parameter setting, I pause to take inventory of the key features of the constant parameter SVAR that make inference for $(\mathbf{A}, \mathbf{F})$ straightforward. The key conceptual properties are:

P1. Observational equivalence of points $(\mathbf{A}, \mathbf{F})$ and $(\mathbf{AQ}, \mathbf{FQ})$ for $\mathbf{Q} \in \mathcal{O}_n$ in the structural model.

P2. The ability to reparameterize the structural model from $(\mathbf{A}, \mathbf{F})$ to $(\mathbf{B}, \mathbf{H}, \mathbf{Q})$ and separate identifiable elements $(\mathbf{B}, \mathbf{H})$ and nonidentifiable elements $\mathbf{Q}$.

From a practical perspective, an additional key property researchers lean on is the following.

P3. The ability to sample the marginal posterior of $(\mathbf{B}, \mathbf{H})$ of the unrestricted structural model as the posterior of a tractable reduced-form model.

In the next section I describe the analogous notions of these properties in a time-varying parameter setting. In section 5 I provide an SVAR with time-varying parameters that delivers the analogous notions of P1, P2, and P3.

3. The Challenge of a Time-Varying Parameter Extension

I take P1, P2, and P3 as the outputs from applying the "want operator" to the notion of a time-varying parameter extension. However, formulating a model that has time-varying parameters, and for which inference can proceed along lines similar to the constant parameter model, poses unique challenges. I next articulate the general class of models under consideration, and then explain the unique issues involved in delivering P1, P2, and P3 in a time-varying parameter setting.

Class of Models Under Consideration. Allowing the structural matrices to change over time, the analogue to equation (2) in the time-varying parameter setting is

$$\mathbf{y}_t'\mathbf{A}_t = \mathbf{x}_t'\mathbf{F}_t + \boldsymbol{\varepsilon}_t', \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n,1}, \mathbf{I}_n), \quad \text{for } 1 \leq t \leq T. \tag{24}$$

From the form of equation (24), one can see that, period-by-period, the data density depends only on $(\mathbf{A}_t, \mathbf{F}_t)$, and thus the joint density of $\mathbf{y}_{1:T}$ takes the form

$$p(\mathbf{y}_{1:T}|\mathbf{A}_{1:T}, \mathbf{F}_{1:T}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{A}_t, \mathbf{F}_t). \tag{25}$$

Period-by-period, $p(\mathbf{y}_t|\mathbf{A}_t, \mathbf{F}_t)$ takes the same form as in the constant-parameter model.

To close the model, one must specify a law of motion for $\mathbf{A}_t$ and $\mathbf{F}_t$. I assume that the law of motion is Markovian and governed by some static parameters $\boldsymbol{\phi}$ taking values in a space $\boldsymbol{\Phi}$. The density of $(\mathbf{A}_t, \mathbf{F}_t)$ then takes the form

$$p(\mathbf{A}_t, \mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}) \tag{26}$$

and

$$p(\mathbf{A}_{1:T}, \mathbf{F}_{1:T}|\boldsymbol{\phi}, \mathbf{A}_0, \mathbf{F}_0) = \prod_{t=1}^{T} p(\mathbf{A}_t, \mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}). \tag{27}$$

To conduct inference in the model, one also needs to specify a distribution for the initial conditions $p(\mathbf{A}_0, \mathbf{F}_0|\boldsymbol{\phi})$. To refer to the density of the entire sequence of time-varying parameters inclusive of the distribution of the initial conditions, I then write

$$p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi}) = p(\mathbf{A}_0, \mathbf{F}_0|\boldsymbol{\phi}) \cdot p(\mathbf{A}_{1:T}, \mathbf{F}_{1:T}|\boldsymbol{\phi}, \mathbf{A}_0, \mathbf{F}_0). \tag{28}$$

For each t, the space of values for the random matrices $(\mathbf{A}_t, \mathbf{F}_t)$ is $\mathbb{S}_t = \mathbb{S}$. As a slight abuse of notation I write $\mathbb{S}^{T+1}$ to refer to the space of values for all of the random matrices in $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$. Each $\mathbb{S}_t = \mathbb{S}$ and hence $\mathbb{S}^{T+1} = \mathbb{S}_0 \times \cdots \times \mathbb{S}_T$. Lastly, one must specify a prior for $\boldsymbol{\phi}$. Having fully specified the Bayesian model, the posterior for all of the model's unobservables takes the form

$$p(\boldsymbol{\phi}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \mathbf{y}_{1:T}) \propto p(\boldsymbol{\phi}) \cdot p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi}) \cdot p(\mathbf{y}_{1:T} | \mathbf{A}_{1:T}, \mathbf{F}_{1:T}). \tag{29}$$

Lastly, note that a model of this form can be interpreted as state-space model, where the unobserved latent states $(\mathbf{A}_t, \mathbf{F}_t)$ are related to the data via the "measurement" equation in (24). I will often adopt this terminology going forward.

3.1 P1 in a Structural TVP model

From equation (29) one can see that inference for $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ will depend on two terms. First, the term $p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi})$, which represents the ability of the particular probability model for the time-varying parameters to rationalize a given sequence $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$. Second, the term $p(\mathbf{y}_{1:T} | \mathbf{A}_{1:T}, \mathbf{F}_{1:T})$, which represents the ability of $(\mathbf{A}_{1:T}, \mathbf{F}_{1:T})$ to rationalize the data. The first of these terms essentially plays the role of a prior for the full sequence $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$, albeit a prior that happens to be conditional on $\boldsymbol{\phi}$ and to be constructed from the particular form of equations (26), (27), and (28).

The observation that $p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi})$ always plays a role in likelihood-based inference for the model's unobservables is the key element of my notion of observational equivalence of parameter points in the time-varying parameter model, which I base on the density

$$p(\mathbf{y}_{1:T}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi}) = p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi}) \cdot p(\mathbf{y}_{1:T} | \mathbf{A}_{1:T}, \mathbf{F}_{1:T}) \tag{30}$$

and formalize in the following definition.

Definition 2 (Observational equivalence)**.** *The points* $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ *and* $(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T})$ *are observationally equivalent if, for any* $\boldsymbol{\phi} \in \boldsymbol{\Phi}$*,*

$$p(\mathbf{y}_{1:T}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi}) = p(\mathbf{y}_{1:T}, \widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T} | \boldsymbol{\phi}) \,.$$

The intuition behind Definition 2 is as follows. If the model of time variation inherently distinguishes between two points $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ and $(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T})$ by assigning them different densities under $p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \boldsymbol{\phi})$, then any likelihood-based procedure for inference will also distinguish between them. Since the differing densities under the law of motion will pass through to equation (30), it would be erroneous to consider such points observationally equivalent.

From the period-by-period symmetry between equation (24) in the TVP case and equation (2) in the constant-parameter case, it is apparent that, regardless of the details of the law of motion, the following claim holds for the invariance of $p(\mathbf{y}_{1:T}|\mathbf{A}_{1:T}, \mathbf{F}_{1:T})$ to orthogonal rotations.

Proposition 1. *If the points* $(\mathbf{A}_{1:T}, \mathbf{F}_{1:T})$ *and* $(\widetilde{\mathbf{A}}_{1:T}, \widetilde{\mathbf{F}}_{1:T})$ *are such that, in each* t, $(\widetilde{\mathbf{A}}_t, \widetilde{\mathbf{F}}_t) = (\mathbf{A}_t\mathbf{Q}_t, \mathbf{F}_t\mathbf{Q}_t)$ *for some* $\mathbf{Q}_t \in \mathcal{O}_n$*, then*

$$p(\mathbf{y}_{1:T}|\mathbf{A}_{1:T}, \mathbf{F}_{1:T}) = p(\mathbf{y}_{1:T}|\widetilde{\mathbf{A}}_{1:T}, \widetilde{\mathbf{F}}_{1:T})\,.$$

For the proof, see Appendix A.1.

Based on Proposition 1 alone, one might have been tempted to conclude that, regardless of the other details of the model, as long as equation (24) is the model's measurement equation, all orthogonally rotated sequences of parameter points are observationally equivalent. But this is not so. Likelihood-based inference about $(\boldsymbol{\phi}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ will depend on the joint density

$$p(\mathbf{y}_{1:T}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi}) = p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi}) \cdot p(\mathbf{y}_{1:T}|\mathbf{A}_{1:T}, \mathbf{F}_{1:T})\,, \tag{31}$$

which includes not only the data density addressed by Proposition 1, but also the model of time variation $p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi})$. Hence, if the points $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ are to be observationally equivalent under orthogonal rotations, then, in addition to Proposition 1, the model of time variation will need to be such that $p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi}) = p(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T}|\boldsymbol{\phi})$ for any $\boldsymbol{\phi}$.

3.2 P2 in a Structural TVP model

If a structural TVP model is such that points differing by orthogonal rotations are observationally equivalent according to Definition 2, then, as in the constant-parameter case, one might hope to reparameterize the model to separate out the nonidentifiable elements.

One can reparameterize the random matrices $(\mathbf{A}_t, \mathbf{F}_t)$ in each t via $(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t) = f(\mathbf{A}_t, \mathbf{F}_t)$, just as in the constant-parameter case, and doing so gives the measurement equation

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B}_t + \boldsymbol{\varepsilon}_t'\mathbf{Q}_t'\underline{h}(\mathbf{H}_t)^{-1'} \tag{32}$$

and induces a law of motion in the new parameter space of the form

$$p(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t | \boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1})\,. \tag{33}$$

Similarly to the previous section, I write $\mathbb{DO}^{T+1}$ to refer to the space containing the random matrices $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T})$.

Regardless of the specifics of the law of motion, the data density conditional on $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T})$ will take the form of equation (7) period-by-period and hence can be written without recourse to $\mathbf{Q}_{0:T}$ as

$$p(\mathbf{y}_{1:T} | \mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}) = p(\mathbf{y}_{1:T} | \mathbf{B}_{0:T}, \mathbf{H}_{0:T})\,. \tag{34}$$

The posterior then takes the form

$$\begin{aligned} p(\mathbf{B}_{0:T},&\mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}) \\ &\propto p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi}) \cdot p(\mathbf{y}_{1:T} | \mathbf{B}_{0:T}, \mathbf{H}_{0:T})\,. \end{aligned} \tag{35}$$

For inference to proceed along similar lines as in the constant-parameter case, one must then factor the density $p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi})$ as

$$p(\mathbf{Q}_{0:T} | \boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \cdot p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi})\,. \tag{36}$$

so that the posterior can be written as

$$\begin{aligned} p(\mathbf{B}_{0:T},&\mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}) \\ &= p(\mathbf{Q}_{0:T} | \boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \cdot p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T})\,, \end{aligned} \tag{37}$$

analogous to equation (14) for the constant-parameter model.

However, attempts to work with the densities in equations (36) and (37) face unique challenges not present in the constant-parameter case. The first challenge is simply obtaining the factorization of equation (36) from the structural law of motion. To see why this is not straightforward, note that, on a period-by-period basis, one may generally factor equation (33) as

$$p(\mathbf{Q}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1}, \mathbf{B}_t, \mathbf{H}_t) \cdot p(\mathbf{B}_t, \mathbf{H}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1}). \tag{38}$$

At first, equation (38) appears analogous to the prior factorization in equation (9) in the constant-parameter case. Note, however, that in general the distribution of $(\mathbf{B}_t, \mathbf{H}_t)$ remains dependent on the particular structural parameters in time $t-1$ through $\mathbf{Q}_{t-1}$. The presence of the $\mathbf{Q}_{t-1}$ in the law of motion for $(\mathbf{B}_t, \mathbf{H}_t)$ impedes the construction of the factorization in equation (33). Hence, the structural model's law of motion needs to yield densities, under reparameterization by f, that decompose as

$$p(\mathbf{Q}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1}, \mathbf{B}_t, \mathbf{H}_t) \cdot p(\mathbf{B}_t, \mathbf{H}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}), \tag{39}$$

where the key thing to note is the absence of $\mathbf{Q}_{t-1}$ from the conditioning information for the density of $(\mathbf{B}_t, \mathbf{H}_t)$. If the law of motion can be decomposed as in equation (39) rather than just (38), then the density for the whole sequence can be written as in equation (36). Finally, this will yield a posterior of the form in equation (37).

3.3 P3 in a Structural TVP model

Lastly, and as a practical matter, the density $p(\mathbf{B}_t, \mathbf{H}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1})$ needs to be tractable enough to conduct Bayesian inference in a purely reduced-form model consisting of

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B}_t + \mathbf{u}_t', \quad \mathbf{u}_t' \sim N(\mathbf{0}_{n,1}, \mathbf{H}_t^{-1}), \quad \text{for } 1 \leq t \leq T \tag{40}$$

$$(\mathbf{B}_t, \mathbf{H}_t) \sim p(\mathbf{B}_t, \mathbf{H}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}). \tag{41}$$

It must also be possible to sample from the conditional prior/posterior density $p(\mathbf{Q}_{0:T}|\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ in equation (37) to complete the task of structural inference. With these properties in place, structural inference can proceed with an algorithm analogous to Algorithm 1.

4. The Extant Approach (and Its Shortcomings)

The approach currently used in the literature is based on estimating the time-varying parameter VAR of Primiceri (2005), which takes the form

$$\mathbf{y}_t' = \mathrm{vec}(\mathbf{B}_t)'(\mathbf{I}_n \otimes \mathbf{x}_t) + \boldsymbol{\varepsilon}_t' \boldsymbol{\Xi}_t \boldsymbol{\Delta}_t^{-1} \quad \text{for} \quad \boldsymbol{\varepsilon}_t \sim N\left(\mathbf{0}_{n,1}, \mathbf{I}_n\right) \tag{42}$$

where

$$\boldsymbol{\Xi}_t = \begin{bmatrix} \xi_{1,t} & 0 & \cdots & 0 \\ 0 & \xi_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \xi_{n,t} \end{bmatrix} \quad , \quad \boldsymbol{\Delta}_t = \begin{bmatrix} 1 & \delta_{12,t} & \cdots & \delta_{1n,t} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta_{n-1n,t} \\ 0 & \cdots & 0 & 1 \end{bmatrix} . \tag{43}$$

Let $\boldsymbol{\delta}_t$ be a vector containing the elements of $\boldsymbol{\Delta}_t$ that are not restricted to be 0 or 1 and define $\boldsymbol{\xi}_t \equiv [\xi_{1,t}, \dots, \xi_{n,t}]$ so that $\boldsymbol{\Xi}_t = \mathrm{diag}(\boldsymbol{\xi}_t)$. The time-varying parameters $(\boldsymbol{\xi}_t, \boldsymbol{\delta}_t, \mathbf{B}_t)$ evolve according to

$$\boldsymbol{\Xi}_t = \boldsymbol{\Xi}_{t-1}\, \mathrm{diag}(\exp(\boldsymbol{\eta}_t)) \ , \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}_{n,1}, \boldsymbol{\Sigma}_{\eta}) \tag{44}$$

$$\boldsymbol{\delta}_t = \boldsymbol{\delta}_{t-1} + \boldsymbol{\zeta}_t \ , \qquad \boldsymbol{\zeta}_t \sim N(\mathbf{0}_{\frac{n(n-1)}{2},1}, \boldsymbol{\Sigma}_{\zeta}) \tag{45}$$

$$\mathrm{vec}(\mathbf{B}_t) = \mathrm{vec}(\mathbf{B}_{t-1}) + \boldsymbol{\upsilon}_t \ , \qquad \boldsymbol{\upsilon}_t \sim N(\mathbf{0}_{mn,1}, \boldsymbol{\Sigma}_{\upsilon}) . \tag{46}$$

The static parameters of the model then consist of $\boldsymbol{\phi} \equiv (\boldsymbol{\Sigma}_{\eta}, \boldsymbol{\Sigma}_{\zeta}, \boldsymbol{\Sigma}_{\upsilon})$.

Researchers call this model a reduced form and then make inference by sampling from a distribution over orthogonal rotations of the parameters. In particular, they operationalize the algorithm I summarize in Algorithm 2, ostensibly in the spirit of Algorithm 1.

Algorithm 2 - Structural posterior sampling with Primiceri (2005) as reduced form

---

1. ***Reduced-form model estimation.***[11]

- *For* $i = 1, \dots, n_1$

[11] One can estimate the parameters of this model using the MCMC algorithm in either Baumeister and Peersman (2013b) or Del Negro and Primiceri (2015).

(a) *sample* $\left(\boldsymbol{\xi}^{(i)}_{0:T}, \boldsymbol{\delta}^{(i)}_{0:T}, \mathbf{B}^{(i)}_{0:T}, \boldsymbol{\phi}^{(i)}\right) \sim p(\boldsymbol{\xi}_{0:T}, \boldsymbol{\delta}_{0:T}, \mathbf{B}_{0:T}, \boldsymbol{\phi}|\mathbf{y}_{1:T})$

(b) *for each t, construct* $\boldsymbol{\Sigma}^{(i)}_{\boldsymbol{\Delta},\boldsymbol{\Xi},t} = (\boldsymbol{\Delta}^{(i)}_t)^{-1\prime}(\boldsymbol{\Xi}^{(i)}_t)^2(\boldsymbol{\Delta}^{(i)}_t)^{-1}$.

- *Store* $\left\{\mathbf{B}^{(i)}_{0:T}, \boldsymbol{\Sigma}^{(i)}_{\boldsymbol{\Delta},\boldsymbol{\Xi},0:T}\right\}_{i=1}^{n_1}$ *and call them the reduced-form parameters.*

2. ***Structural inference.***

- *For each draw in* $\left\{\mathbf{B}^{(i)}_{0:T}, \boldsymbol{\Sigma}^{(i)}_{\boldsymbol{\Delta},\boldsymbol{\Xi},0:T}\right\}_{i=1}^{n_1}$, *sample* $\mathbf{Q}^{(i)}_{0:T} \sim p^{\mathcal{R}}(\mathbf{Q}_{0:T}|\mathbf{B}^{(i)}_{0:T}, \boldsymbol{\Sigma}^{(i)}_{\boldsymbol{\Delta},\boldsymbol{\Xi},0:T})$

In the context of the discussion in Section 3, one might say this approach starts from the pieces of the decomposition in equations (36) and (39) in the discussion of P2, rather than deriving them as results from densities for $(\mathbf{A}_t, \mathbf{F}_t)$ that satisfy P1. In the context of Algorithm 1 for the constant-parameter case, the second step implements a method for choosing among observationally equivalent parameters in the structural representation. In contrast, Algorithm 2 proceeds without having formulated a structural representation that rationalizes all of the candidate parameters sampled in the second step as observationally equivalent.

Of course, when taking such an approach, it is also the case that all properties of the model formulated over $(\mathbf{B}_t, \mathbf{H}_t)$ will affect inference for the structural parameters in the second step. When using Primiceri (2005) as the model for the evolution of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$, one such known property is the dependence of estimates of $(\boldsymbol{\xi}_{0:T}, \boldsymbol{\delta}_{0:T}, \mathbf{B}_{0:T}, \boldsymbol{\phi})$ on the particular ordering of the variables in $\mathbf{y}_t$ during estimation. In other words, when using Primiceri (2005) as a device for eliciting inference on $(\boldsymbol{\Sigma}^{(i)}_{\boldsymbol{\Delta},\boldsymbol{\Xi},0:T}, \mathbf{B}_{0:T})$, as in Algorithm 2, an n variable system admits $n!$ different answers to any question asked of the model.[12] To see this property, consider the distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},t}$ under the laws of motion in equations (44) and (45) in a 2-variable example.[13] In the 2-variable case,

$$\text{(47)} \qquad \boldsymbol{\Delta}_t = \begin{bmatrix} 1 & \delta_{12,t} \\ 0 & 1 \end{bmatrix} \qquad \Longleftrightarrow \qquad \boldsymbol{\Delta}_t^{-1} = \begin{bmatrix} 1 & -\delta_{12,t} \\ 0 & 1 \end{bmatrix}$$

---

[12] Note that estimating even a single specification of the model is computationally demanding. A single specification of a VAR-TVP-SV with four variables and four lags can easily take 24 hours to estimate and another 24 hours for structural inference via the application of sign restrictions to the first stage's estimation output. Factorials behaving as they do, even a medium-sized system of 20 variables admits 2.43×10^{18} different orderings. In the absence of analytical results about the estimator's properties under alternative orderings, statements about the significance of variable ordering in such a setting are obviously pure conjecture.

[13] This simple example is mentioned in footnote 5 of Primiceri (2005).

and hence

$$\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},t} = \boldsymbol{\Delta}_t^{-1\prime}\boldsymbol{\Xi}_t\boldsymbol{\Xi}_t\boldsymbol{\Delta}_t^{-1} = \begin{bmatrix} \xi_{1,t}^2 & -\delta_{12,t}\xi_{1,t}^2 \\ -\delta_{12,t}\xi_{1,t}^2 & \delta_{12,t}^2\xi_{1,t}^2 + \xi_{2,t}^2 \end{bmatrix}. \tag{48}$$

Substituting for the elements of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},t}$ in terms of the stochastic processes defined by equations (43)–(45), the conditional distributions of the diagonal elements are given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[1,1],t} = \xi_{1,t}^2 = (\xi_{1,t-1}\underbrace{\exp(\boldsymbol{\eta}_{1,t})}_{\sim\,\text{Lognormal}})^2 \sim \text{Lognormal} \tag{49}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[2,2],t} &= \delta_{12,t}^2\xi_{1,t}^2 + \xi_{2,t}^2 \\ &= \underbrace{(\delta_{12,t-1} + \boldsymbol{\zeta}_{1,t})^2}_{\sim\,\text{scaled noncentral }\chi^2}\ \underbrace{(\xi_{1,t-1}\exp(\boldsymbol{\eta}_{1,t}))^2}_{\sim\,\text{Lognormal}} + \underbrace{(\xi_{2,t-1}\exp(\boldsymbol{\eta}_{2,t}))^2}_{\sim\,\text{Lognormal}} \\ &\nsim \text{Lognormal} \end{aligned} \tag{50}$$

Thus the distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[1,1],t}$ is lognormal, while the distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[2,2],t}$ is not. Flipping the order of variables 1 and 2 prior to estimation flips each variable's pertinent element on the diagonal of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},t}$. Since the distributions of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[1,1],t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[2,2],t}$ belong to different distributional families, the density of each variable's respective element of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},t}$ necessarily changes.[14] When inference under alternative orderings leads to different inference for $(\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},0:T}, \mathbf{B}_{0:T})$, it then also feeds into different inference for $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$. I next show that this property can have significant consequences in practice for structural inference.

An Example Application. Consider the application of Baumeister and Peersman (2013b) to the identification of the time-varying effects of oil supply shocks. In step 1 of Algorithm 2 the authors estimate the model of Primiceri (2005) with

[14] Another popular version of this model is that of Cogley and Sargent (2005), in which $\boldsymbol{\Xi}_t$ is time-varying but $\boldsymbol{\Delta}_t$ is constant. Restricting $\boldsymbol{\Delta}_t$ to be constant still yields random variables $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[1,1],t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[2,2],t}$ belonging to different distributional families and thus the same issue is present. To see this, note that the distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[1,1],t}$ in equation (49) is unaffected while, in equation (50), the leading term in deriving the distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[2,2],t}$ becomes a constant instead of a scaled noncentral χ^2 random variable. Nonetheless, the distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta},\boldsymbol{\Xi},[2,2],t}$ still involves the sum of lognormal random variables, which does not yield a lognormal random variable.

$n = 4$ observables ordered as

$$\mathbf{y}_t' = [\Delta q_t^{oil}, \Delta p_t^{oil}, \Delta gdp_t, \Delta cpi_t]. \tag{51}$$

Having collected the draws of $\left\{\mathbf{\Sigma}_{\mathbf{\Delta},\mathbf{\Xi},0:T}, \mathbf{B}_{0:T}\right\}$, Baumeister and Peersman (2013b) proceed to structural inference in step 2. The identifying assumptions for the oil supply shock are that a negative oil supply shock causes, contemporaneously and for four quarters thereafter, the quantity of oil supplied to fall and the price of oil to increase.[15]

Baumeister and Peersman (2013b) find an increasing response over time of p^{oil} to a shock that causes q^{oil} to fall by 1 percent and they generally argue that the oil demand curve has steepened over time. Ostensibly, the quantitative results follow from the relatively uncontroversial sign restrictions for identifying the oil supply shocks. However, to the extent that the empirical results were meant to follow primarily from the sign restrictions, the dynamic system could just as easily have been estimated under $4! = 24$ different orderings of the variables in $\mathbf{y}_t$.

It turns out that alternative orderings of $\mathbf{y}_t$ can give substantively different inference for the structural quantities of interest. Figure 1 shows the (generalized) IRFs under the paper's original ordering (left column of plots in Figure 1), as well as the results from two alternative variable orderings (right column of plots in Figure 1). In all cases the results are generated by Algorithm 2 and have identical sign-restrictions.[16] The first alternative ordering reverses the ordering of all variables, in which case one can see that the maximum impulse responses are roughly half of what they are under the baseline results, and there is significantly

[15]To be precise, Baumeister and Peersman (2013b) apply the sign restrictions to generalized impulse responses. The requirement for the oil supply shock at time t, denoted $\varepsilon_t^{oil,s}$, is that in response to $\varepsilon_t^{oil,s} < 0$,

$$E\left[\sum_{h=0}^{\bar{h}} \Delta q_{t+h}^{oil}\right] < 0 < E\left[\sum_{h=0}^{\bar{h}} \Delta p_{t+h}^{oil}\right] \tag{52}$$

for each of $\bar{h} = 0, \ldots, 4$.

[16]I estimate each model using the same specification as Baumeister and Peersman (2013b) for the MCMC algorithm: 50,000 iterations of "burn-in," followed by 50,000 iterations for estimation, from which every 10th draw is retained, thus yielding 5,000 posterior draws used for inference.

less evidence of meaningful time variation in the impulse responses. The second alternative ordering reverses the order of only the second and third variables, in which case the maximum responses are roughly twice as large as they are under the baseline ordering.[17] With no guidance on which ordering is "more correct" than another, any of the results presented in Figure 1 could have been considered the "main results" *with no aspect of the method pointing to them as more wrong or right than any of the others.*

To be clear, this example does not say that Baumeister and Peersman (2013b) are "wrong" about the structural question *per se*. However, it does highlight a deeply problematic property of pursuing set identification of structural quantities based on using the Primiceri (2005) model as a device for formulating a law of motion for the VAR coefficients.

5. The Drifting SVAR

In this section I describe my new structural VAR with time-varying parameters that delivers P1, P2, and P3. To emphasize the parallels to the consant-parameter framework, the organization of this section mirrors that of Section 2.

5.1 DSVAR: The Structural Model with Time-Varying Parameters

To more easily reference the full model, I repeat the measurement equation given in Section 3,

$$\mathbf{y}_t'\mathbf{A}_t = \mathbf{x}_t'\mathbf{F}_t + \boldsymbol{\varepsilon}_t', \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n,1}, \mathbf{I}_n), \quad \text{for } 1 \leq t \leq T. \tag{24}$$

The key novelty of my approach is to specify the laws of motion for $\mathbf{A}_t$ and $\mathbf{F}_t$, which define $p(\mathbf{A}_t, \mathbf{F}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1})$, as

$$\mathbf{A}_t = \beta^{-1/2}\,\mathbf{A}_{t-1}\boldsymbol{\Omega}_t \qquad \text{for} \qquad \boldsymbol{\Omega}_t \sim p(\boldsymbol{\Omega}_t) \tag{53}$$

$$\mathbf{F}_t = \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t + \boldsymbol{\Theta}_t \qquad \text{for} \qquad \boldsymbol{\Theta}_t \sim N(\mathbf{0}_{m,n}, \mathbf{W}, \mathbf{I}_n) \tag{54}$$

where β is a scalar, $\boldsymbol{\Omega}_t$ is an $(n \times n)$ matrix of random shocks that *multiplicatively* perturb $\mathbf{A}_{t-1}$, and $\boldsymbol{\Theta}_t$ is an $m \times n$ random matrix of matrix-normal mean-zero

[17] The data for this exercise come from the replication files for Baumeister and Peersman (2013b), which are publicly available at `https://www.aeaweb.org/articles?id=10.1257/mac.5.4.1`.

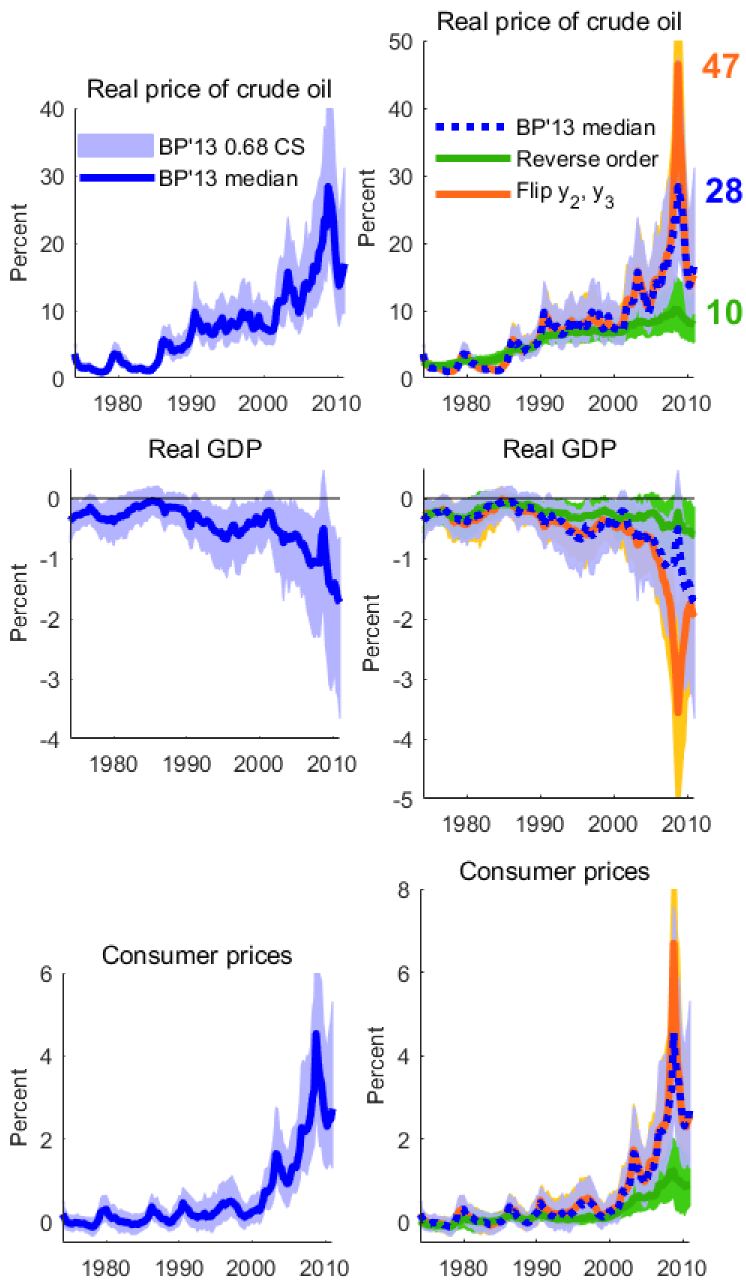

FIGURE 1.—Impulse responses of macroeconomic variables to an oil supply shock causing a 1 percent decrease in world oil production. Numbers in right margin indicate maximum of a model's time-series of median responses.

additive shocks, where $\mathbf{W}$ is an $(m \times m)$ symmetric positive definite matrix. The laws of motion and distributions of shocks in equations (53) and (54) in turn pin down $p(\mathbf{A}_t, \mathbf{F}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1})$ in the model. In this model $\boldsymbol{\phi} \equiv (\beta, \mathbf{W})$. For the moment I remain agnostic about the distribution of $\boldsymbol{\Omega}_t$ beyond the requirement that it satisfies the following condition.

Condition 1 (Spherical multiplicative shocks)**.** *For any* $\boldsymbol{\Omega}_t$ *and any* $\mathbf{Q}, \mathbf{P} \in \mathcal{O}_n$, $p(\boldsymbol{\Omega}_t) = p(\mathbf{Q}\boldsymbol{\Omega}_t\mathbf{P})$.

Condition 1 requires that the density of $\boldsymbol{\Omega}_t$ is invariant to multiplication from the left and right by (possibly different) orthogonal matrices, a property for which the standard term in multivariate statistics is "spherical." Note that there are many candidate distributions for $\boldsymbol{\Omega}_t$ that satisfy Condition 1, for example $p(\boldsymbol{\Omega}_t) \sim N(\mathbf{0}_{n\times n}, \mathbf{I}_n, \mathbf{I}_n)$ would do the trick. I refer to a model with laws of motion in the form of equations (24), (53), and (54) and with $p(\boldsymbol{\Omega}_t)$ satisfying Condition 1 as a *drifting* SVAR, or DSVAR for short.

The notion of structural shocks now includes two types of disturbances. The first type consists of the vector of shocks $\boldsymbol{\varepsilon}_t$, which perturb $\mathbf{y}_t$ through the equilibrium dynamics represented by $(\mathbf{A}_t, \mathbf{F}_t)$, and which also appeared in the constant-parameter SVAR. The realization of $\boldsymbol{\varepsilon}_t$ affects $\mathbf{y}_t$ and $\mathbf{y}_{t+1}$ and so on through the VAR's dependence on lagged values, but it does not affect the structural parameters of the system and hence does not alter impulse response functions. The second type consists of the random matrices $(\boldsymbol{\Omega}_t, \boldsymbol{\Theta}_t)$, which perturb the coefficients governing the equilibrium relationships among the variables, thus altering impulse responses.

The key properties of the DSVAR go through largely from the structure already described. However, because the distribution for initial conditions will always play a role in inference, the following condition will also be required to complete some of the subsequent arguments.

Condition 2 (Orthogonal invariance of distribution of initial conditions)**.** *For any* $\mathbf{Q}_0 \in \mathcal{O}_n$ *and any* $\boldsymbol{\phi} \in \boldsymbol{\Phi}$*, the density for the initial conditions* $(\mathbf{A}_0, \mathbf{F}_0)$ *satisfies* $p\big(\mathbf{A}_0, \mathbf{F}_0 | \boldsymbol{\phi}\big) = p\big(\mathbf{A}_0\mathbf{Q}_0, \mathbf{F}_0\mathbf{Q}_0 | \boldsymbol{\phi}\big)$.

5.2 The Identification Problem in a DSVAR

Turning to the density of $(\mathbf{A}_{1:T}, \mathbf{F}_{1:T})$ under the model's law of motion, the following result for the density of sequences of structural parameters is obtained.

Theorem 1. *In a DSVAR, if the points* $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ *and* $(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T})$ *are such that, in each t,* $(\widetilde{\mathbf{A}}_t, \widetilde{\mathbf{F}}_t) = (\mathbf{A}_t\mathbf{Q}_t, \mathbf{F}_t\mathbf{Q}_t)$ *for some* $\mathbf{Q}_t \in \mathcal{O}_n$*, then*

$$p(\mathbf{A}_{1:T}, \mathbf{F}_{1:T}|\boldsymbol{\phi}, \mathbf{A}_0, \mathbf{F}_0) = p(\widetilde{\mathbf{A}}_{1:T}, \widetilde{\mathbf{F}}_{1:T}|\boldsymbol{\phi}, \widetilde{\mathbf{A}}_0, \widetilde{\mathbf{F}}_0)\,.$$

If, additionally, $p\big(\mathbf{A}_0, \mathbf{F}_0|\boldsymbol{\phi}\big)$ *satisfies Condition 2, then*

$$p(\mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi}) = p(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T}|\boldsymbol{\phi})\,.$$

For the proof, see Appendix A.2.

Theorem 1 says that for any parameter point of the structural model, all other parameter points that differ by orthogonal rotations have the same density. As should be apparent from the discussion in Section 3.1, this is critical for establishing observational equivalence under orthogonal rotations, but it is also a surprising result. For example, the rotation of a point $(\mathbf{A}_t, \mathbf{F}_t)$ into $(\widetilde{\mathbf{A}}_t, \widetilde{\mathbf{F}}_t)$ must be reconciled with two "moving parts" since both the density of $\mathbf{A}_t, \mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}$ and the density of $\mathbf{A}_{t+1}, \mathbf{F}_{t+1}|\mathbf{A}_t, \mathbf{F}_t$ are affected. With the additional condition on the distribution of initial conditions, the statement applies to full sequences of the structural parameters.

From the orthogonal invariance of the law of motion shown in Theorem 1, and the orthogonal invariance of the conditional data density shown in Proposition 1, the ingredients are in place for a statement about observational equivalence in a DSVAR.

Theorem 2. *In a DSVAR, if the points* $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ *and* $(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T})$ *are related as in Theorem 1,*

$$p(\mathbf{y}_{1:T}, \mathbf{A}_{1:T}, \mathbf{F}_{1:T}|\boldsymbol{\phi}, \mathbf{A}_0, \mathbf{F}_0) = p(\mathbf{y}_{1:T}, \widetilde{\mathbf{A}}_{1:T}, \widetilde{\mathbf{F}}_{1:T}|\boldsymbol{\phi}, \widetilde{\mathbf{A}}_0, \widetilde{\mathbf{F}}_0)\,.$$

If, additionally, $p\big(\mathbf{A}_0, \mathbf{F}_0|\boldsymbol{\phi}\big)$ *satisfies Condition 2, then*

$$p(\mathbf{y}_{1:T}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T}|\boldsymbol{\phi}) = p(\mathbf{y}_{1:T}, \widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T}|\boldsymbol{\phi})\,.$$

Thus, the points $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ *and* $(\widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T})$ *are observationally equivalent under Definition 2.*

Proof. Examining the factorization in equation (31), the result follows directly from applying Proposition 1 to the conditional data density and Theorem 1 to the density of the dynamic parameters. □

The same conditions yield an analogous statement for the posterior of all unobservables.

Corollary 1. *Under the conditions of Theorem 2, the posterior of the unobservables satisfies* $p(\boldsymbol{\phi}, \mathbf{A}_{0:T}, \mathbf{F}_{0:T} | \mathbf{y}_{1:T}) = p(\boldsymbol{\phi}, \widetilde{\mathbf{A}}_{0:T}, \widetilde{\mathbf{F}}_{0:T} | \mathbf{y}_{1:T})$.

Proof. Examining the factorization in equation (29), the result follows immediately from Theorem 2. □

Theorem 2 (and Corollary 1) mean that for any realization of $\mathbf{y}_{1:T}$, the econometrician cannot differentiate between parameter points that differ by orthogonal rotations.

5.3 A Useful Reparameterization of a DSVAR

The DSVAR's invariance to orthogonal rotations of parameters suggests the potential utility of reparameterizing in terms of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T})$, similar to the constant-parameter model.

First, consider making this reparameterization for the structural parameters in a single period t. The joint density is given by

$$\begin{aligned} p\big(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}\big) = {} & p_{\mathbf{A}_t, \mathbf{F}_t}\big(f^{-1}(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t) | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}\big) \\ & \cdot J\big((\mathbf{A}_t, \mathbf{F}_t) \longrightarrow (\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t)\big) \end{aligned} \tag{55}$$

The following result summarizes the implications of the reparameterization of the time t structural parameters.

Proposition 2. *In a DSVAR*

(i) $p\big(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}\big) = p(\mathbf{Q}_t) \cdot p\big(\mathbf{B}_t, \mathbf{H}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}\big)$

(ii) $p\big(\mathbf{A}_{t+1}, \mathbf{F}_{t+1} | \boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t\big) = p\big(\mathbf{A}_{t+1}, \mathbf{F}_{t+1} | \boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_t\big)$.

For the proof, see the appendix.

The key elements of Proposition 2 to note are the independence of $\mathbf{Q}_t$ and $(\mathbf{B}_t, \mathbf{H}_t)$ in (i) and the lack of dependence of $(\mathbf{A}_{t+1}, \mathbf{F}_{t+1})$ on $\mathbf{Q}_t$ in (ii).

Proposition 3. *In a DSVAR,*

$$p\big(\mathbf{B}_{t+1}, \mathbf{H}_{t+1}, \mathbf{Q}_{t+1} | \boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t\big) = p(\mathbf{Q}_{t+1}) \cdot p\big(\mathbf{B}_{t+1}, \mathbf{H}_{t+1}, | \boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_t\big).$$

For the proof, see the appendix.

As a result of the properties summarized in Proposition 3, reparameterizing all elements of the sequence $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ gives

Corollary 2. *In a DSVAR, if the distribution for initial conditions satisfies Condition 2, then*

$$p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi}\big) = p\big(\mathbf{Q}_{0:T}\big) \cdot p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}\big) \tag{56}$$

where

$$p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}\big) = p\big(\mathbf{B}_0, \mathbf{H}_0 | \boldsymbol{\phi}\big) \cdot \prod_{t=1}^{T} p\big(\mathbf{B}_t, \mathbf{H}_t | \boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}\big) \tag{57}$$

$$p\big(\mathbf{Q}_{0:T}\big) = \prod_{t=0}^{T} p(\mathbf{Q}_t). \tag{58}$$

Proof. The proof follows from the fact that $p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi}\big) = p\big(\mathbf{B}_0, \mathbf{H}_0, \mathbf{Q}_0 | \boldsymbol{\phi}\big) \cdot \prod_{t=1}^{T} p\big(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t | \boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1})$ and then applying Proposition 3 in each t. □

The statement about the law of motion in Corollary 2 leads directly to the following result for the posterior density conditional on $\boldsymbol{\phi}$.

Corollary 3. *In a DSVAR, if the distribution for initial conditions satisfies Condition 2, then*

$$p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}\big) = p(\mathbf{Q}_{0:T}) \cdot p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T})$$

Proof. Given the factorization in Corollary 2, the result follows directly from equations (35), (36), and (37). □

Lastly, one can find the following result for the joint posterior of all unobservables.

Corollary 4. *In a DSVAR, if the distribution for initial conditions satisfies Condition 2, then*

$$p\big(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\mathbf{y}_{1:T}\big) = p(\mathbf{Q}_{0:T}) \cdot p(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\mathbf{y}_{1:T})$$

Proof. Noting that

$$p\big(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\mathbf{y}_{1:T}\big) \propto p(\boldsymbol{\phi}) \cdot p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\boldsymbol{\phi}, \mathbf{y}_{1:T}\big), \tag{59}$$

the result follows from Corollary 3. □

When analyzing the reduced form of a DSVAR it will be useful to note the following additional properties of the densities in Proposition 3.

Proposition 4. *The densities in Proposition 3 take the forms:*

$$p(\mathbf{Q}_{t+1}) = U(\mathcal{O}_n) \tag{60}$$

$$p\big(\mathbf{B}_{t+1}, \mathbf{H}_{t+1}, |\boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_t\big) = p(\mathbf{H}_{t+1}|\boldsymbol{\phi}, \mathbf{H}_t) \cdot p(\mathbf{B}_{t+1}|\boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_{t+1}) \tag{61}$$

where

$$p\big(\mathbf{H}_{t+1}, |\boldsymbol{\phi}, \mathbf{H}_t\big) \propto p_{\boldsymbol{\Omega}}\big(\beta^{1/2}\underline{h}(\mathbf{H}_{t-1})^{-1}\underline{h}(\mathbf{H}_t)\big) \cdot \beta^{n^2/2}|\underline{h}(\mathbf{H}_{t-1})|^{-n}|\mathbf{H}_t|^{-1/2} \tag{62}$$

$$p(\mathbf{B}_{t+1}|\boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_{t+1}) = N(\mathbf{B}_{t+1}|\mathbf{B}_t, \mathbf{W}, \mathbf{H}_{t+1}^{-1})\,. \tag{63}$$

For the proof, see the appendix.

5.4 Identifying Restrictions in a DSVAR

As in the constant-parameter model, the economist will typically impose restrictions on the set of candidates for structural parameters by restricting attention to certain regions of the parameter space. In general, one might represent such restrictions as a set $\mathcal{R} \subseteq \mathbb{D}\mathbb{O}^{T+1}$. To take a simple example, one could impose some standard restrictions from the SVAR literature on a period-by-period basis. Note, however, that the restriction region now pertains to the entire joint space of all random matrices in $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T})$, so more exotic identification schemes are possible.

Analagous to equation (15) in the constant-parameter model, one can consider the prior in the restricted model to be

$$
\begin{aligned}
&p^{\mathcal{R}}\big(\mathbf{B}_{0:T},\mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\boldsymbol{\phi}\big) \\
&\qquad = p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\boldsymbol{\phi}\big) \cdot I\big\{\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}\big) \in \mathcal{R}\big\}.
\end{aligned} \tag{64}
$$

which can in general be factored into a marginal and conditional similarly to the unrestricted case, as

$$
\begin{aligned}
&p^{\mathcal{R}}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\boldsymbol{\phi}) \\
&\qquad \propto\ p(\mathbf{Q}_{0:T}|\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \cdot I\big\{\mathbf{Q}_{0:T} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})\big\} \\
&\qquad\quad \cdot p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\boldsymbol{\phi}) \cdot I\big\{(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \in \mathcal{R}_{\mathbf{B,H}}\big\}
\end{aligned} \tag{65}
$$

where

$$
\mathcal{R}_{\mathbf{B,H}} \equiv \big\{(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) : \exists\, \mathbf{Q}_{0:T} \text{ for which } (\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}) \in \mathcal{R}\big\} \tag{66}
$$

$$
\mathcal{R}_{\mathbf{Q}}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \equiv \big\{\mathbf{Q}_{0:T} : \big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}\big) \in \mathcal{R}\big\}. \tag{67}
$$

Under the assumption that the identifying restrictions are, again, "exact or less," in the sense that $I\big\{(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \in \mathcal{R}_{\mathbf{B,H}}\big\} = 1$ for almost all $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \in \mathbb{D}_{T+1}$, I define the restricted prior for $\mathbf{Q}_{0:T}$ as

$$
\begin{aligned}
&p^{\mathcal{R}}(\mathbf{Q}_{0:T}|\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \\
&\qquad = \frac{p(\mathbf{Q}_{0:T}|\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \cdot I\big\{\mathbf{Q}_{0:T} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})\big\}}{\int_{\mathbf{Q}_{0:T}} p(\mathbf{Q}_{0:T}|\mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \cdot I\big\{\mathbf{Q}_{0:T} \in \mathcal{R}_{\mathbf{Q}}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})\big\} d\mathbf{Q}_{0:T}}.
\end{aligned} \tag{68}
$$

The posterior of the restricted model then becomes

$$
\begin{aligned}
&p^{\mathcal{R}}\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}|\mathbf{y}_{1:T}, \boldsymbol{\phi}\big) \\
&\qquad = p^{\mathcal{R}}(\mathbf{Q}_{0:T}|\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \cdot p\big(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\mathbf{y}_{1:T}, \boldsymbol{\phi}\big)\,.
\end{aligned} \tag{69}
$$

5.5 Reduced Form of a DSVAR

In this section I show that the structural model of the previous section admits a set of reduced-form parameters as a marginal for which one can construct a purely reduced-form representation.

5.5.1 Fully Reduced Form in General

From the factorization in Corollary 3, one can see that, in principle, a DSVAR should admit a reduced-form representation for the dynamics of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$. Under the reparameterization of the previous section, the basic idea is to integrate out the unobservable matrices $\mathbf{Q}_{0:T}$ from the posterior. The sought-after fully reduced-form model then needs to specify laws of motion for $(\mathbf{B}_t, \mathbf{H}_t)$ such that the posterior of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ is equivalent to the marginal posterior under the structural model. To do so, it will suffice to construct laws of motion for $(\mathbf{B}_t, \mathbf{H}_t)$ that induce the DSVAR's density for $p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\boldsymbol{\phi})$, which is how I proceed.

Proposition 5. *Let $p(\mathbf{\Omega}_t)$ be as defined for a DSVAR and let $p_{\mathbf{\Omega}\mathbf{\Omega}'}$ denote the density of the random matrix $\mathbf{\Omega}_t\mathbf{\Omega}_t'$. If $\mathbf{H}_t$ evolves according to the law of motion*

$$\mathbf{H}_t = \frac{1}{\beta}\,\underline{h}(\mathbf{H}_{t-1})\,\mathbf{\Gamma}_t\,\underline{h}(\mathbf{H}_{t-1})' \quad \textit{for} \quad \mathbf{\Gamma}_t \sim p_{\mathbf{\Omega}\mathbf{\Omega}'}(\mathbf{\Gamma}_t)$$

then the distribution $p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1})$ is the same as that given in Proposition 4.

For the proof, see the appendix.

Next, turning to the law of motion for $\mathbf{B}_t$,

Proposition 6. *The distribution $p(\mathbf{B}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_t)$ induced by the law of motion*

$$\mathbf{B}_t = \mathbf{B}_{t-1} + \mathbf{V}_t \quad \textit{for} \quad \mathbf{V}_t \sim \mathrm{N}\left(\mathbf{0}_{m,n}, \mathbf{W}, \mathbf{H}_t^{-1}\right),$$

is the same as that given in Proposition 4.

Proof. From Proposition 4, the requisite distribution for $\mathbf{B}_t$ is $N(\mathbf{B}_t|\mathbf{B}_{t-1}, \mathbf{W}, \mathbf{H}_t^{-1})$. The result then follows from the form of $\mathbf{V}_t$ and well-known properties of the matrix-variate normal distribution under affine transformations. □

The upshot of the previous two results is that, in principle, one can obtain the marginal posterior of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ by estimating a reduced-form model directly. I codify this fact in the following two results.

Theorem 3. *Let $p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi})$ be the density from a reparameterized DSVAR, as given in Proposition 3, and let $\widetilde{p}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi})$ be the density under the purely-reduced-form model, as implied by Proposition 5 and Proposition 6. Then*

$$p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}) = \widetilde{p}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}) .$$

Proof. The proof follows directly from Proposition 5 and Proposition 6. □

Which leads to the following statement about the Bayesian posteriors conditional on $\boldsymbol{\phi}$.

Corollary 5. *In a DSVAR, with the $p(\cdot)$ and $\widetilde{p}(\cdot)$ notation defined analogously to Theorem 3,*

$$p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}) = \widetilde{p}(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}) .$$

Proof. Noting that

$$p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}) \propto p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}) \cdot p(\mathbf{y}_{1:T} | \mathbf{B}_{0:T}, \mathbf{H}_{0:T}), \tag{70}$$

the result follows from Theorem 3. □

Lastly, I get the following result for the Bayesian posteriors of the unobservables including $\boldsymbol{\phi}$.

Corollary 6. *In a DSVAR, with the $p(\cdot)$ and $\widetilde{p}(\cdot)$ notation defined analogously to Theorem 3,*

$$p(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}, | \mathbf{y}_{1:T}) = \widetilde{p}(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}, | \mathbf{y}_{1:T}) .$$

Proof. Note that

$$p(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \mathbf{y}_{1:T}) \propto p(\boldsymbol{\phi}) \cdot p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}) \cdot p(\mathbf{y}_{1:T} | \mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \tag{71}$$

$$\propto p(\boldsymbol{\phi}) \cdot p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \boldsymbol{\phi}, \mathbf{y}_{1:T}). \tag{72}$$

The result then follows from Corollary 5. □

5.5.2 A Tractable Special Case

The model that I take to the data in the subsequent sections uses a distribution for $\mathbf{\Omega}_t$ that induces a matrix beta distribution for $\mathbf{\Gamma}_t = \mathbf{\Omega}_t\mathbf{\Omega}_t'$, denoted $B_n(\nu_1/2, \nu_2/2)$. To do so, I define a density for the shocks $\mathbf{\Omega}_t$ that is essentially a random "matrix square root" of a $B_n(\nu_1/2, \nu_2/2)$-distributed matrix. I leave the details on the construction of this random matrix for Appendix A.4. For present purposes, the key thing to note is that inducing $\mathbf{\Gamma}_t \sim B_n(\nu_1/2, \nu_2/2)$ has the key benefit of yielding a reduced-form model, known in the statistics literature as a dynamic linear model with discounted Wishart stochastic volatility (DLM-DWSV), that is tractable and that nests the constant-parameter VAR as a limiting case. The most tractable form of the DLM-DWSV, and the one that I take to the data in subsequent sections, uses $\nu_1 = \beta/(1-\beta)$ and $\nu_2 = 1$.[18]

Two key model features allow the DLM-DWSV to nest the VAR. First, one can collapse to $\mathbf{0}_{m,n}$ the distribution for the additive shocks to $\mathbf{B}_t$ by taking $\mathbf{W} \to \mathbf{0}_{m,m}$. With no shocks perturbing the coefficients in $\mathbf{B}_t$ from one period to the next, each $\mathbf{B}_t$ is pulled to $\mathbf{B}_0$, eliminating the model's time-varying VAR coefficients. $\mathbf{B}_0$ then plays the role of $\mathbf{B}$ in the constant-parameter case. Second, under certain choices of ν_1 and ν_2 in which they are linked to β, taking $\beta \to 1$ pulls $\mathbf{\Gamma}_t \to \beta\mathbf{I}_n$ for all t.[19] From the law of motion in Proposition 5, one can see that forcing $\mathbf{\Gamma}_t = \beta\mathbf{I}_n$ will have the effect of pulling each $\mathbf{H}_t$ to $\mathbf{H}_{t-1}$ and so on all the way back to $\mathbf{H}_0$, eliminating the model's stochastic volatility of forecast errors. $\mathbf{H}_0$ then plays the role of $\mathbf{H}$ in the constant-parameter case.

5.6 From Reduced-Form Parameter Estimation to Structural Inference in a DSVAR

Given inference for $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \boldsymbol{\phi})$, inference for $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}, \boldsymbol{\phi})$ subject to identifying restrictions can proceed in much the same way as in constant-parameter models. Analogous to Algorithm 1, researchers can generate samples of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T}, \boldsymbol{\phi})$ by implementing Algorithm 3.

---

[18] With $\nu_2 = 1$, the distribution is known as singular matrix beta with density derived by Uhlig (1994).

[19] When the shocks $\mathbf{\Gamma}_t$ are distributed as $B_n(\nu_1/2, \nu_2/2)$, they have $\mathrm{E}[\mathbf{\Gamma}_t] = \nu_1/(\nu_1+\nu_2)\,\mathbf{I}_n$. Thus, setting $\nu_1/\nu_2 = \beta/(1-\beta)$ gives $\mathrm{E}[\mathbf{\Gamma}_t] = \beta\,\mathbf{I}_n$. Since $\mathrm{E}[\mathbf{H}_t|\beta, \mathbf{H}_{t-1}] = \beta^{-1}\mathbf{H}_{t-1}^{1/2}\mathrm{E}[\mathbf{\Gamma}_t]\mathbf{H}_{t-1}^{1/2'}$, the process for $\mathbf{H}_t$ given in Proposition 5 then takes on random walk behavior, i.e., $\mathrm{E}[\mathbf{H}_t|\beta, \mathbf{H}_{t-1}] = \mathbf{H}_{t-1}$, when $\mathrm{E}[\mathbf{\Gamma}_t] = \beta\,\mathbf{I}_n$. For the result on the collapsing second moments, see Konno (1988).

Algorithm 3 - Structural posterior sampling for a DSVAR

1. ***Reduced-form model estimation***

   - *For* $i = 1, \dots, n_1$, *sample* $\left(\mathbf{B}_{0:T}^{(i)}, \mathbf{H}_{0:T}^{(i)}\right) \sim p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\mathbf{y}_{1:T})$
   - *Store* $\left\{\mathbf{B}_{0:T}^{(i)}, \mathbf{H}_{0:T}^{(i)}\right\}_{i=1}^{n_1}$

2. ***Structural inference***

   - *For each draw in* $\left\{\mathbf{B}_{0:T}^{(i)}, \mathbf{H}_{0:T}^{(i)}\right\}_{i=1}^{n_1}$, *sample* $\mathbf{Q}_{0:T}^{(i)} \sim p^{\mathcal{R}}(\mathbf{Q}_{0:T}|\mathbf{B}_{0:T}^{(i)}, \mathbf{H}_{0:T}^{(i)})$.

5.7 The DSVAR's Tractability (or The "Wants" for a TVP Extension Delivered)

Before moving on to the empirical application, I pause to take inventory of the key features of the DSVAR that make inference for $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ straightforward. The key conceptual properties are:

P1. Observational equivalence of points $(\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ and $(\mathbf{AQ}, \mathbf{FQ})$ for $\mathbf{Q}_{0:T} \in \mathcal{O}_n$ in the structural model.

P2. The ability to reparameterize the structural model to $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T}, \mathbf{Q}_{0:T})$ to separate identifiable elements $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ and nonidentifiable elements $\mathbf{Q}_{0:T}$.

From a practical perspective, an additional key property researchers require is the following.

P3. The DSVAR is described in enough generality here that this will vary across model formulations; however, Section 5.5.2 gives a special case that is particularly tractable (described in greater detail in the next section).

6. Bayesian Estimation of the Reduced-Form Model with Matrix-Beta Shocks

Given the data $\mathbf{y}_{1:T}$, Bayesian estimation of the reduced-form model entails characterizing the posterior distribution of the model's unobservables:

$$p(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \mathbf{y}_{1:T}) = \frac{p(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})\, p(\mathbf{y}_{1:T} | \boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})}{p(\mathbf{y}_{1:T})} \tag{73}$$

After having estimated $(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})$, inference for structural parameters can proceed as described in Section 5.6.

One cannot fully characterize the posterior in equation (73) analytically so I make inference about $(\boldsymbol{\phi}, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ by generating a random sample from the posterior via a Markov chain Monte Carlo (MCMC) algorithm. MCMC algorithms iterate over a Markov chain constructed to have the posterior distribution as its invariant distribution. While draws from the MCMC algorithm are not iid, iteratively sampling from the MCMC algorithm asymptotically yields draws representative of the model's posterior. In particular, my MCMC algorithm is of a type known as a Gibbs sampler, which means that the algorithm entails iteratively sampling from the *conditional* posteriors of the subsets of a partition of the model's unobservables.

My Gibbs sampler for the DLM-DWSV consists of two blocks of parameters based on the partition of the unobservables into $\mathbf{W}$ and $(\beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})$. The goal then becomes sampling from the conditional distributions,

1. **Block 1.** $p(\mathbf{W} | \mathbf{y}_{1:T}, \beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T})$
2. **Block 2.** $p(\beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \mathbf{y}_{1:T}, \mathbf{W})$
   (a) $p(\beta | \mathbf{y}_{1:T}, \mathbf{W})$
   (b) $p(\mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \mathbf{y}_{1:T}, \beta, \mathbf{W})$

I leave the exact formulas for Appendix B, but the high-level considerations to be mindful of are as follows. Under an inverse Wishart prior for $\mathbf{W}$, which may condition on the value of β, the conditional posterior in Block 1 is also an inverse Wishart distribution; hence, this draw is straightforward. In Block 2 the sample from the joint posterior of $\beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T} | \mathbf{y}_{1:T}, \mathbf{W}$ is achieved by factoring the joint distribution into the distribution of $\beta|\mathbf{y}_{1:T}, \mathbf{W}$ in Step 2(a), which is *marginal* of $\mathbf{B}_{0:T}, \mathbf{H}_{0:T}$, and the distribution of $\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\mathbf{y}_{1:T}, \mathbf{W}, \beta$ in Step 2(b), which conditions on the value of β.

My sampling strategy for Block 2 makes use of two particularly elegant properties of the DLM-DWSV. First, there exist exact expressions for evaluating the likelihood for the static parameters marginal of the entire sequence $\mathbf{B}_{0:T}, \mathbf{H}_{0:T}$, *including the stochastic volatility components*, in a fashion analogous to likelihood-based inference with the Kalman filter. I describe the steps of the recursive filter in Table I.[20] The draw of $\beta|\mathbf{y}_{1:T}, \mathbf{W}$ in Step 2(a) can then be

TABLE I

RECURSIVE FILTER FOR LIKELIHOOD EVALUATION IN THE DLM-DWSV

Distribution of Interest	Distributional Family	Parameters	Supporting Computations
Step 1 – Prior			
		$\left(d_{t-1\|t-1}, \boldsymbol{\Psi}_{t-1\|t-1}, \overline{\mathbf{B}}_{t-1\|t-1}, \mathbf{C}_{t-1\|t-1}\right)$ given from iteration $t-1$	
$(\mathbf{H}_t\|\mathbf{y}_{1:t-1}, \boldsymbol{\phi})$	$W\left(d_{t\|t-1}, \boldsymbol{\Psi}_{t\|t-1}^{-1}\right)$	$d_{t\|t-1} = \beta d_{t-1\|t-1}$ $\boldsymbol{\Psi}_{t\|t-1} = \beta \boldsymbol{\Psi}_{t-1\|t-1}$	
$(\mathbf{B}_t\|\mathbf{y}_{1:t-1}, \boldsymbol{\phi}, \mathbf{H}_t)$	$N\left(\overline{\mathbf{B}}_{t\|t-1}, \mathbf{C}_{t\|t-1}, \mathbf{H}_t^{-1}\right)$	$\overline{\mathbf{B}}_{t\|t-1} = \overline{\mathbf{B}}_{t-1\|t-1}$ $\mathbf{C}_{t\|t-1} = \mathbf{C}_{t-1\|t-1} + \mathbf{W}$	
Step 1.5 – Prediction			
$(\mathbf{y}_t\|\mathbf{y}_{1:t-1}, \boldsymbol{\phi})$	$T_{\zeta_t}\left(\overline{\mathbf{y}}_{t\|t-1}, \boldsymbol{\Sigma}_{\mathbf{y}_t}\right)$	$\zeta_t = d_{t\|t-1} - n + 1$ $\overline{\mathbf{y}}_{t\|t-1} = \overline{\mathbf{B}}'_{t\|t-1}\mathbf{x}_t$ $\boldsymbol{\Sigma}_{\mathbf{y}_t} = (q_t/\zeta_t)\boldsymbol{\Psi}_{t\|t-1}$	 $q_t = \mathbf{x}_t'\mathbf{C}_{t\|t-1}\mathbf{x}_t + 1$
Step 2 – Posterior			
$(\mathbf{H}_t\|\mathbf{y}_{1:t}, \boldsymbol{\phi})$	$W\left(d_{t\|t}, \boldsymbol{\Psi}_{t\|t}^{-1}\right)$	$d_{t\|t} = d_{t\|t-1} + 1$ $\boldsymbol{\Psi}_{t\|t} = \boldsymbol{\Psi}_{t\|t-1} + \frac{1}{q_t}\mathbf{e}_t\mathbf{e}_t'$	 $\mathbf{e}_t = \mathbf{y}_t - \overline{\mathbf{y}}_{t\|t-1}$
$(\mathbf{B}_t\|\mathbf{y}_{1:t}, \boldsymbol{\phi}, \mathbf{H}_t)$	$N\left(\overline{\mathbf{B}}_{t\|t}, \mathbf{C}_{t\|t}, \mathbf{H}_t^{-1}\right)$	$\overline{\mathbf{B}}_{t\|t} = \overline{\mathbf{B}}_{t\|t-1} + \mathbf{K}_t\mathbf{e}_t'$ $\mathbf{C}_{t\|t} = \mathbf{C}_{t\|t-1} - \mathbf{K}_t\mathbf{K}_t'q_t$	$\mathbf{K}_t = \mathbf{C}_{t\|t-1}\mathbf{x}_t/q_t$

implemented with a Metropolis-Hastings step since one can evaluate (a kernel

[20]Table I summarizes the relevant results from the statistics literature; see Prado and West (2010).

of) the target density pointwise. Evaluating the log-likelihood function pointwise amounts to simply running the algorithm forward and accumulating the sum of evaluations of the log of the density in Step 1.5 at the realized values of $\mathbf{y}_t$. Second, step 2(b) can be implemented via known exact expressions for recursively sampling backwards the sequence of latent states. I summarize the algorithm for this draw, sometimes referred to as a "simulation smoother," in Table II.[21]

TABLE II
SIMULATION SMOOTHER FOR DLM-DWSV

Distribution to Be Sampled	Distributional Family	Parameters and Supporting Computations
		$\left(d_{t\mid t}, \mathbf{\Psi}_{t\mid t}, \overline{\mathbf{B}}_{t\mid t}, \mathbf{C}_{t\mid t}, \overline{\mathbf{B}}_{t+1\mid t}, \mathbf{C}_{t+1\mid t}\right)$ given from forwards filter
$(\mathbf{H}_t \mid \mathbf{y}_{1:T}, \boldsymbol{\phi}, \mathbf{H}_{t+1})$	$\mathbf{H}_t = \beta \mathbf{H}_{t+1} + \mathbf{\Upsilon}_t$ $\mathbf{\Upsilon}_t \sim W(d^*_{t\mid t+1}, \mathbf{\Psi}^{-1}_{t\mid t})$	$d^*_{t\mid t+1} = (1-\beta) d_{t\mid t}$
$(\mathbf{B}_t \mid \mathbf{y}_{1:T}, \boldsymbol{\phi}, \mathbf{H}_t, \mathbf{B}_{t+1})$	$N(\overline{\mathbf{B}}_{t\mid t+1}, \mathbf{C}_{t\mid t+1}, \mathbf{H}^{-1}_t)$	$\widetilde{\mathbf{K}}_t = \mathbf{C}_{t\mid t}\mathbf{C}^{-1}_{t+1\mid t}$ $\overline{\mathbf{B}}_{t\mid t+1} = \overline{\mathbf{B}}_{t\mid t} + \widetilde{\mathbf{K}}_t(\mathbf{B}_{t+1} - \overline{\mathbf{B}}_{t+1\mid t})$ $\mathbf{C}_{t\mid t+1} = \mathbf{C}_{t\mid t} - \widetilde{\mathbf{K}}_t \mathbf{C}_{t+1\mid t} \widetilde{\mathbf{K}}'_t$

6.1 Priors

The model primitives requiring prior distributions are β,$\mathbf{W}$, and $(\mathbf{B}_0, \mathbf{H}_0)$. This section describes the general structure of my priors, with the specific choices of prior hyperparameters provided in the context of the application.

Prior for β. The uncertainty over $\mathbf{H}_t$ at each step of the filter is characterized by a Wishart distribution with degrees of freedom of either βh_{t-1} or $\beta h_{t-1} + 1$ (see Table I). When starting the h_t values at their steady state of $1/(1-\beta)$, the smaller of these two degrees of freedom parameters is $\beta/(1-\beta)$. To maintain, at each step of the filter, a valid Wishart probability distribution and a valid multivariate-*t*

[21] The distribution of $\mathbf{B}_t$ given in Table II corrects an erratum in Prado and West (2010) (M. West, personal communication, January 23, 2018). The corrected version is publicly available by incorporating the changes in `http://www2.stat.duke.edu/~mw/Prado&WestBook/errata.pdf`.

predictive density for $\mathbf{y}_t$, β then needs to satisfy $\beta/(1-\beta) > n-1$ and thus $p(\beta)$ can have positive density only on $((n-1)/n, 1)$. I use a 4-parameter beta distribution, which allows one to set the min and max values in addition to the usual shape and scale parameters of the beta distribution.

Prior for W. As described in equation (46), in the model of Primiceri (2005) the distribution of the model's linear coefficients takes the form $\text{vec}(\mathbf{B}_t) \sim N(\text{vec}(\mathbf{B}_{t-1}), \boldsymbol{\Sigma}_v)$. It has become standard in the literature to base the prior for $\boldsymbol{\Sigma}_v$ on a presample of observations. In particular, a standard choice is

$$\boldsymbol{\Sigma}_v \sim IW(T_{pre},\ k^2_{\Sigma_v} \cdot T_{pre} \cdot V(\hat{B}_{OLS})) \tag{74}$$

where T_{pre} is the number of pre-sample observations, k_Q is a hyperparameter chosen by the researcher, and $V(\hat{b}_{OLS})$ is the matrix of standard errors for the $\hat{b}_{pre}$ OLS estimates.[22] In Primiceri (2005), $k_Q = 0.01$ and $T_{pre} = 40$ and $V(\hat{B}_{OLS}) = \Sigma_{pre} \otimes (X'_{pre}X_{pre})^{-1}$.

In the DLM-DWSV, $\text{vec}(\mathbf{B}_t)$ has distribution

$$\text{vec}(\mathbf{B}_t) \sim N(\text{vec}(\mathbf{B}_{t-1}), \boldsymbol{\Sigma}_t \otimes \mathbf{W}) \tag{75}$$

and hence the matrix $(\mathbf{H}_t^{-1} \otimes \mathbf{W})$ functions similarly to Q from the Primiceri (2005) model. I choose a prior for $\mathbf{W}$ in the spirit of the $p(Q)$ given in (74). Estimating a VAR over a presample under a diffuse prior yields the posterior for $\mathbf{b}_{pre}$ of

$$\mathbf{b}_{pre}|\Sigma_{pre} \sim N(\hat{b}_{pre}, \Sigma_{pre} \otimes (X'_{pre}X_{pre})^{-1})\,. \tag{76}$$

I then scale the prior according to the number of presample observations and a hyperparameter δ_1^2.

$$\mathbf{W} \sim IW(\delta_1^2 \cdot T_{pre} \cdot (X'_{pre}X_{pre})^{-1}, T_{pre})\,. \tag{77}$$

Prior for $(\mathbf{B}_0, \mathbf{H}_0)$. The prior for the initial values of the dynamic latent states $(\mathbf{B}_0, \mathbf{H}_0)$ maintains the form of the distributional families in the recursive filter

[22] Clark and Ravazzolo (2015) follow this procedure as well.

summarized in Table I, i.e., $(\mathbf{B}_0, \mathbf{H}_0) \sim NW(\overline{\mathbf{B}}_{0|0}, \mathbf{C}_{0|0}, d_{0|0}, \mathbf{\Psi}_{0|0}^{-1})$. The prior for $(\mathbf{B}_1, \mathbf{H}_1)$ is then induced by the model's law of motion and inference for its sufficient statistics is obtained according to Step 1 of Table I. Treating $(\mathbf{B}_0, \mathbf{H}_0)$ in this fashion allows its elements to be integrated out of the likelihood, just like the rest of the sequence $(\mathbf{B}_{1:T}, \mathbf{H}_{1:T})$. The remaining primitives to be specified are then $(\overline{\mathbf{B}}_{0|0}, \mathbf{C}_{0|0}, d_{0|0}, \mathbf{\Psi}_{0|0}^{-1})$. In the context of the application, I leave discussion of the specific choices for these hyperparameters to the appendix.

7. Application: Revisiting a Time-Varying Oil Demand Elasticity

I apply the DSVAR to revisit the extent of time-varying price elasticities in the oil market. The basic set-up is the same as in the motivating example of Section 4. Figure 2 plots the impulse responses from my model plotted against the impulse responses from the 3 orderings in the motivating example. The results from my DSVAR have some similarities to those of Baumeister and Peersman (2013b) under their variable ordering, but my results generally indicate smaller and smoother responses, more in line with the results from the reverse ordering.

To further contextualize the results of the DSVAR, I compare it to the results one would obtain from an approach based on Algorithm 2, but which was "agnostic" about which variable ordering was correct. In particular, I estimate all $n! = 24$ possible orderings and then estimate impulse responses by integrating over the simulated impulse responses from all of the models. It is worth emphasizing that, although I can conduct such an exercise in the context of this particular model, such an approach rapidly becomes computationally infeasible as the number of variables moves beyond $n = 4$. Figure 3 shows the results of the DSVAR against the posterior medians and Bayesian credible sets one would obtain by integrating over the impulse responses of all 24 orderings. Interestingly, the posterior medians from the DSVAR are virtually identical to those obtained by considering estimates from Algorithm 2 when accounting for all 24 orderings. Thus the inference obtained from the DSVAR, which entails estimating only a single model, replicates the inference one would obtain from an approximately agnostic approach that keeps the Primiceri (2005) model at the core of the inference.

8. Conclusion

This paper developed a new SVAR with time-varying parameters, which I call a DSVAR. It is the only such model in the literature to date that allows for internally consistent notions of exact and set identification of structural parameters following the estimation of a reduced form. As a byproduct, unlike other approaches in the literature, the methods developed here are invariant to variable ordering in the vector of observables. I apply the model to the application of Baumeister and Peersman (2013b) and show that my proposed method yields inference almost identical to what one would obtain by accounting for all $n!$ possible orderings of the Baumeister and Peersman (2013b) model.

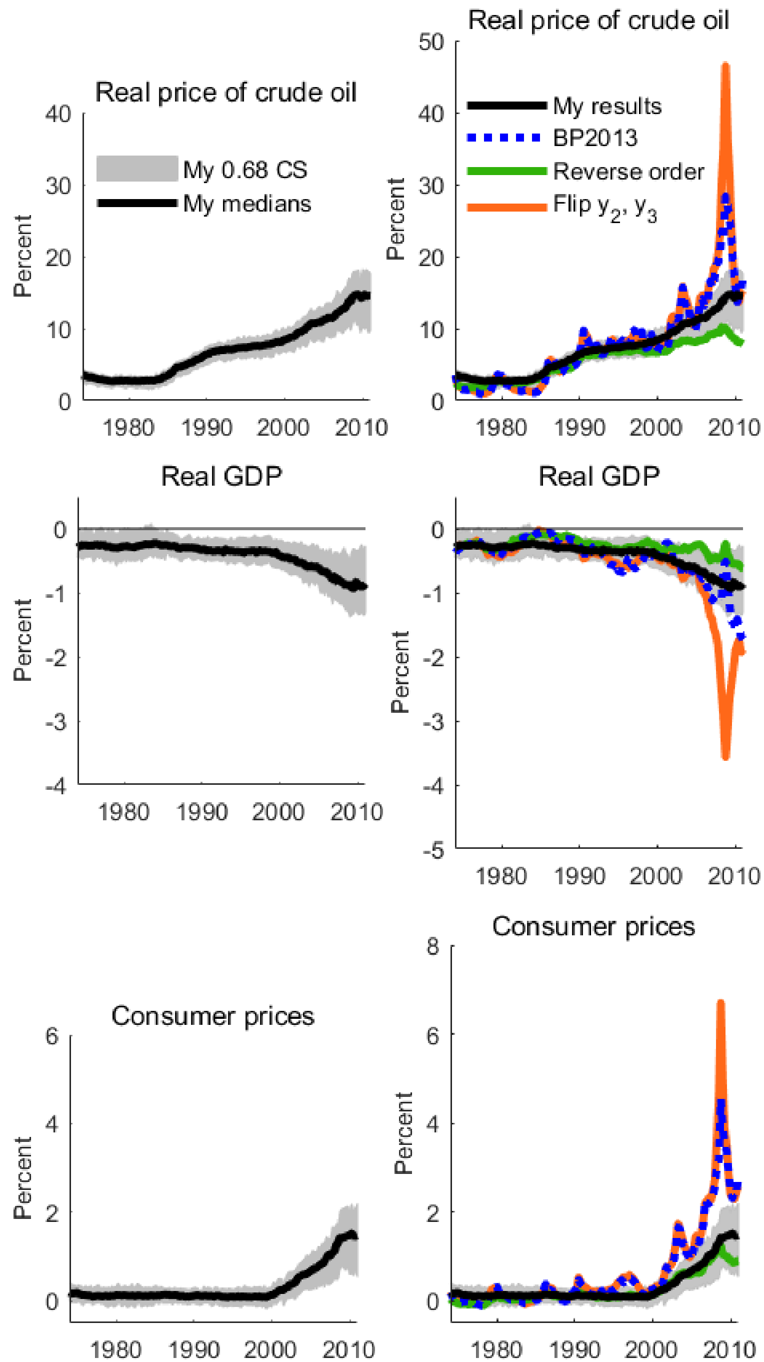

FIGURE 2.—Impulse responses of macroeconomic variables to an oil supply shock causing a 1 percent decrease in world oil production.

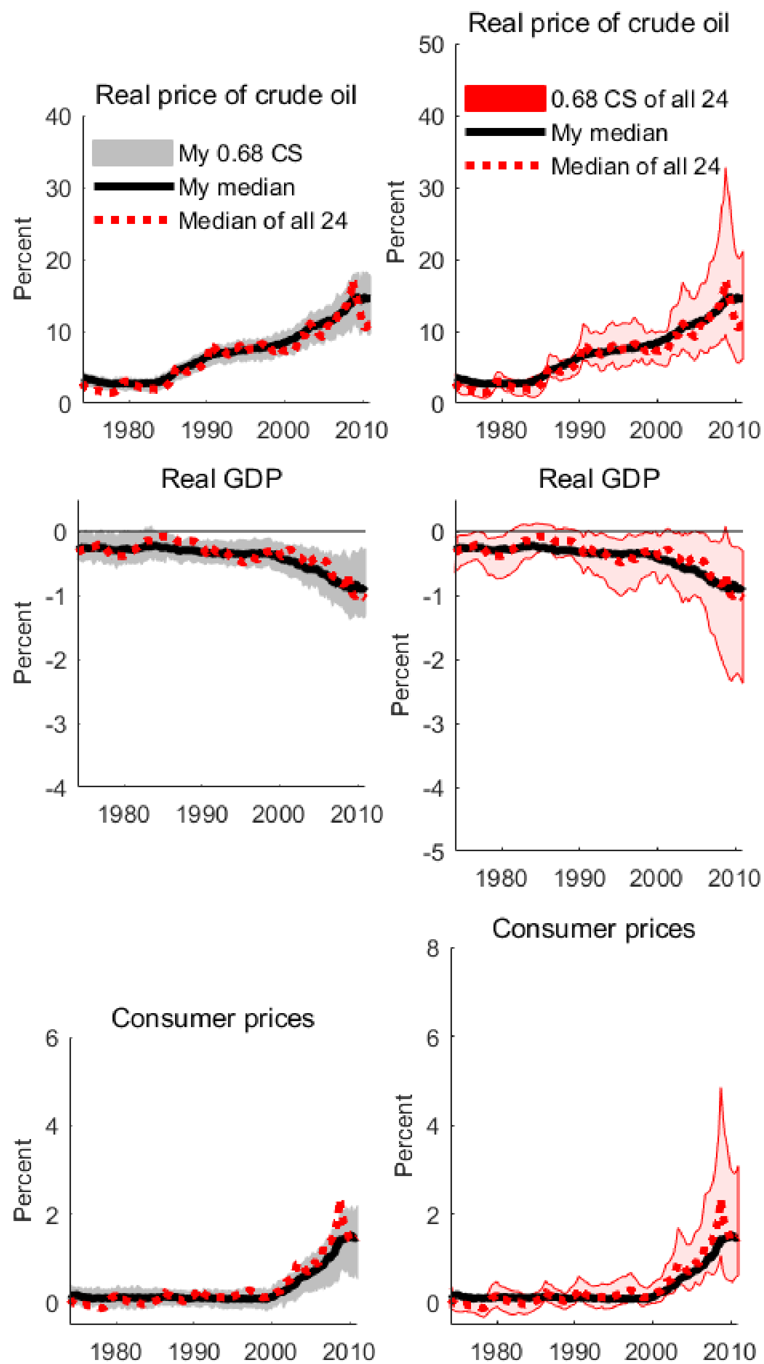

FIGURE 3.—Impulse responses of macroeconomic variables to an oil supply shock causing a 1 percent decrease in world oil production.

A. Proofs

A.1 Useful Results

I first prove two useful results that are well-known in the literature on constant-parameter SVARs, but which will also prove useful in my extension to a TVP model.

Lemma 1. *Let* $(\mathbf{A}, \mathbf{F})$ *be a parameter point in* $\mathbb{S}$*,* $\mathbf{Q}$ *a matrix in* $\mathcal{O}_n$*, and* $(\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}}) = (\mathbf{AQ}, \mathbf{FQ})$*, then* $p(\mathbf{y}_t | \mathbf{A}, \mathbf{F}, \mathbf{y}_{t-p:t-1}) = p(\mathbf{y}_t | \widetilde{\mathbf{A}}, \widetilde{\mathbf{F}}, \mathbf{y}_{t-p:t-1})$.

Proof. First rewrite equation (2) as

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{F}\mathbf{A}^{-1} + \varepsilon_t'\mathbf{A}^{-1}\,. \tag{A.78}$$

Under the assumption that $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n,1}, \mathbf{I}_n)$, the density of $\mathbf{y}_t$ is thus

$$\begin{aligned} p(\mathbf{y}_t | (\mathbf{A}, \mathbf{F}), \mathbf{y}_{t-p:t-1}) &= (2\pi)^{-n/2} |(\mathbf{AA}')^{-1}|^{-1/2} \\ &\quad \cdot \exp\left\{-(1/2) \cdot (\mathbf{y}_t' - \mathbf{x}_t'\mathbf{F}\mathbf{A}^{-1})(\mathbf{AA}')(\mathbf{y}_t' - \mathbf{x}_t'\mathbf{F}\mathbf{A}^{-1})'\right\}\,. \end{aligned} \tag{A.79}$$

The lemma follows from evaluating equation (A.79) at the parameter point $\widetilde{\mathbf{S}}$ and noting that

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}' = \mathbf{AQQ}'\mathbf{A}' = \mathbf{AA}' \tag{A.80}$$

$$\tilde{\mathbf{F}}\tilde{\mathbf{A}}^{-1} = \mathbf{FQQ}^{-1}\mathbf{A}^{-1} = \mathbf{FA}' \tag{A.81}$$

where the second equalities in equations (A.80) and (A.81) follow from the orthogonality of $\mathbf{Q}$. □

The previous result leads to the following corollary for the likelihood of the entire sequence of observables.

Corollary 7. *Let* $(\mathbf{A}, \mathbf{F})$ *and* $(\widetilde{\mathbf{A}}, \widetilde{\mathbf{F}})$ *be as defined in Lemma 1, then* $p(\mathbf{y}_{1:T} | \mathbf{A}, \mathbf{F}) = p(\mathbf{y}_{1:T} | \widetilde{\mathbf{A}}, \widetilde{\mathbf{F}})$.

Proof.

$$p(\mathbf{y}_{1:T}|\mathbf{A},\mathbf{F}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{A},\mathbf{F},\mathbf{y}_{0:t-1}) \tag{A.82}$$

$$= \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{A},\mathbf{F},\mathbf{y}_{t-p:t-1}) \tag{A.83}$$

$$= \prod_{t=1}^{T} p(\mathbf{y}_t|\widetilde{\mathbf{A}},\widetilde{\mathbf{F}},\mathbf{y}_{t-p:t-1}) \tag{A.84}$$

$$= p(\mathbf{y}_{1:T}|\widetilde{\mathbf{A}},\widetilde{\mathbf{F}})\,. \tag{A.85}$$

where the equality in (A.84) follows from Lemma 1. □

These results carry over almost immediately to the time-varying parameter model because each $p(\mathbf{y}_t|\mathbf{A}_t,\mathbf{F}_t,\mathbf{y}_{t-p:t-1})$ takes a form nearly identical to its counterpart in the constant-parameter model.

Corollary 8. *Let* $(\mathbf{A}_t,\mathbf{F}_t)$ *be a parameter point in* $\mathbb{S}_t$*,* $\mathbf{Q}_t$ *a matrix in* $\mathcal{O}_n$*, and* $(\widetilde{\mathbf{A}}_t,\widetilde{\mathbf{F}}_t) = (\mathbf{A}_t\mathbf{Q}_t,\mathbf{F}_t\mathbf{Q}_t)$*, then* $p(\mathbf{y}_t|\mathbf{A}_t,\mathbf{F}_t,\mathbf{y}_{t-p:t-1}) = p(\mathbf{y}_t|\widetilde{\mathbf{A}}_t,\widetilde{\mathbf{F}}_t,\mathbf{y}_{t-p:t-1})$.

Proof. The proof goes through identically to Lemma 1 by adding the subscript t to $\mathbf{A}$, $\mathbf{F}$, and $\mathbf{Q}$. □

This leads to the proof of Proposition 1.

Proof of Proposition 1. The proof goes through identically to Corollary 7 by adding the subscript t to each $\mathbf{S}$ and making recourse to Corollary 8. □

A.2 Proof of Theorem 1

Proof of Theorem 1.
I prove the result in three parts.

Part I: Preliminaries.

The structure of equations (53) and (54) is such that one can factor each conditional density as

$$p(\mathbf{A}_t,\mathbf{F}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) = p(\mathbf{A}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1})\cdot p(\mathbf{F}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{A}_t) \tag{A.86}$$

$$= p(\mathbf{A}_t|\mathbf{A}_{t-1},\boldsymbol{\phi})\cdot p(\mathbf{F}_t|\mathbf{A}_{t-1},\mathbf{F}_{t-1},\boldsymbol{\phi},\mathbf{A}_t) \tag{A.87}$$

where (A.87) follows from the definition of the law of motion for $\mathbf{A}_t$ in equation (53), which does not depend on $\mathbf{F}_{t-1}$. Hence, one can write equation (27) as

$$p(\mathbf{A}_{1:T}, \mathbf{F}_{1:T} | \boldsymbol{\phi}, \mathbf{A}_0, \mathbf{F}_0) = \left[\prod_{t=1}^{T} p(\mathbf{A}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1})\right] \cdot \left[\prod_{t=1}^{T} p(\mathbf{F}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t)\right]. \tag{A.88}$$

To prove the desired result it then suffices to show that, for parameter points related as

$$(\widetilde{\mathbf{A}}_t, \widetilde{\mathbf{F}}_t) = (\mathbf{A}_t \mathbf{Q}_t, \mathbf{F}_t \mathbf{Q}_t) \tag{A.89}$$

$$(\widetilde{\mathbf{A}}_{t-1}, \widetilde{\mathbf{F}}_{t-1}) = (\mathbf{A}_{t-1} \mathbf{Q}_{t-1}, \mathbf{F}_{t-1} \mathbf{Q}_{t-1}), \tag{A.90}$$

the following two equalities hold

$$p(\mathbf{A}_t | \mathbf{A}_{t-1}, \boldsymbol{\phi}) = p(\widetilde{\mathbf{A}}_t | \widetilde{\mathbf{A}}_{t-1}, \boldsymbol{\phi}) \tag{A.91}$$

$$p(\mathbf{F}_t | \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}, \mathbf{A}_t) = p(\widetilde{\mathbf{F}}_t | \widetilde{\mathbf{A}}_{t-1}, \widetilde{\mathbf{F}}_{t-1}, \boldsymbol{\phi}, \widetilde{\mathbf{A}}_t). \tag{A.92}$$

Part II: Proof that equation (A.91) *holds.*

Deriving the conditional density of $\mathbf{A}_t$ as a change of variables from the random matrix $\mathbf{\Omega}_t$ and the law of motion in equation (53) gives

$$p(\mathbf{A}_t | \mathbf{A}_{t-1}, \boldsymbol{\phi}, p_{\mathbf{\Omega}}) = p_{\mathbf{\Omega}}(\beta^{1/2} \mathbf{A}_{t-1}^{-1} \mathbf{A}_t) \cdot J\left(\mathbf{\Omega}_t \longrightarrow \mathbf{A}_t | \mathbf{A}_{t-1}, \boldsymbol{\phi}\right) \tag{A.93}$$

Evaluating the $p_{\mathbf{\Omega}}$ term in equation (A.93) at the point $\widetilde{\mathbf{A}}_t$ and $\widetilde{\mathbf{A}}_{t-1}$ gives

$$p_{\mathbf{\Omega}}(\beta^{1/2} \widetilde{\mathbf{A}}_{t-1}^{-1} \widetilde{\mathbf{A}}_t) = p_{\mathbf{\Omega}}(\beta^{1/2} \mathbf{Q}_{t-1}^{-1} \mathbf{A}_{t-1}^{-1} \mathbf{A}_t \mathbf{Q}_t) \tag{A.94}$$

$$= p_{\mathbf{\Omega}}(\mathbf{Q}_{t-1}^{-1} (\beta^{1/2} \mathbf{A}_{t-1}^{-1} \mathbf{A}_t) \mathbf{Q}_t) \tag{A.95}$$

$$= p_{\mathbf{\Omega}}(\beta^{1/2} \mathbf{A}_{t-1}^{-1} \mathbf{A}_t) \tag{A.96}$$

The equality of the $p_{\mathbf{\Omega}}$ evaluations in equations (A.95) and (A.96) follows from the assumption that $p_{\mathbf{\Omega}}$ satisfies Condition 1 and the fact that both $\mathbf{Q}_{t-1}^{-1}$ and $\mathbf{Q}_t$ are orthogonal matrices.

The Jacobian term is given by

$$J\left(\boldsymbol{\Omega}_t \longrightarrow \mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}\right) = |\beta^{1/2}\mathbf{A}_{t-1}^{-1}|^n = \beta^{n^2/2}|\mathbf{A}_{t-1}|^{-n} \ . \tag{A.97}$$

Evaluated at the point $(\widetilde{\mathbf{A}}_t, \widetilde{\mathbf{A}}_{t-1})$, the Jacobian simplifies as

$$J\left(\boldsymbol{\Omega}_t \longrightarrow \mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}\right)\Big|_{(\mathbf{A}_t,\mathbf{A}_{t-1})=(\widetilde{\mathbf{A}}_t,\widetilde{\mathbf{A}}_{t-1})} = \beta^{n^2/2}|\widetilde{\mathbf{A}}_{t-1}|^{-n} \tag{A.98}$$

$$= \beta^{n^2/2}|\mathbf{A}_{t-1}\mathbf{Q}_{t-1}|^{-n} \tag{A.99}$$

$$= \beta^{n^2/2}|\mathbf{A}_{t-1}|^{-n} \underbrace{|\mathbf{Q}_{t-1}|^{-n}}_{=1} \tag{A.100}$$

$$= \beta^{n^2/2}|\mathbf{A}_{t-1}|^{-n} \tag{A.101}$$

Putting the results in equations (A.96) and (A.101) together, one can see that

$$p(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}, p_{\boldsymbol{\Omega}}) = p(\widetilde{\mathbf{A}}_t|\widetilde{\mathbf{A}}_{t-1}, \boldsymbol{\phi}, p_{\boldsymbol{\Omega}}) \ , \tag{A.102}$$

which completes Part II.

Part III: Proof that equation (A.92) *holds.*

Define $\overline{\mathbf{F}}_t \equiv \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t$. Under the law of motion in equation (54), it is immediate from the definition of $\boldsymbol{\Theta}_t$ and well-known properties of the matrix-variate normal distribution that

$$p(\mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t) = p_N(\mathbf{F}_t|\overline{\mathbf{F}}_t, \mathbf{W}, \mathbf{I}_n) \tag{A.103}$$

Hence, the conditional density of $\mathbf{F}_t$ is given by

$$\begin{aligned} p(\mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}, \mathbf{A}_t) &= (2\pi)^{-nm/2}|\mathbf{I}_n|^{-m/2}|\mathbf{W}|^{-n/2} \\ &\quad \cdot \exp\{-\frac{1}{2}\mathrm{tr}[\mathbf{I}_n^{-1}(\mathbf{F}_t - \overline{\mathbf{F}}_t)'\mathbf{W}^{-1}(\mathbf{F}_t - \overline{\mathbf{F}}_t)]\} \end{aligned} \tag{A.104}$$

Now consider a random matrix $\widehat{\mathbf{F}}_t$ where

$$\widehat{\mathbf{F}}_t = \widetilde{\mathbf{F}}_{t-1}\widetilde{\mathbf{A}}_{t-1}^{-1}\widetilde{\mathbf{A}}_t + \mathbf{\Theta}_t \tag{A.105}$$

$$= \left(\mathbf{F}_{t-1}\mathbf{Q}_{t-1}\right)\left(\mathbf{Q}_{t-1}^{-1}\mathbf{A}_{t-1}^{-1}\right)\left(\mathbf{A}_t\mathbf{Q}_t\right) + \mathbf{\Theta}_t \tag{A.106}$$

$$= \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{Q}_t + \mathbf{\Theta}_t \tag{A.107}$$

$$= \overline{\mathbf{F}}_t\mathbf{Q}_t + \mathbf{\Theta}_t \tag{A.108}$$

and hence

$$\widehat{\mathbf{F}}_t|\widetilde{\mathbf{A}}_{t-1}, \widetilde{\mathbf{F}}_{t-1}, \boldsymbol{\phi}, \widetilde{\mathbf{A}}_t \sim N(\overline{\mathbf{F}}_t\mathbf{Q}_t,\ \mathbf{W},\ \mathbf{I}_n) \tag{A.109}$$

From inspection of the expressions in (A.103) and (A.109), one can see that the density of $\widehat{\mathbf{F}}_t$ and the density of $\mathbf{F}_t$ differ by only their means, which appear only in the exponential-trace term of the matrix-variate normal density. The exponential-trace term from the density of $\widehat{\mathbf{F}}_t$ is

$$\operatorname{tr}\left[(\widehat{\mathbf{F}}_t - \overline{\mathbf{F}}_t\mathbf{Q}_t)'\mathbf{W}^{-1}(\widehat{\mathbf{F}}_t - \overline{\mathbf{F}}\mathbf{Q}_t)\right]. \tag{A.110}$$

Evaluating the expression in (A.110) at the point $\widehat{\mathbf{F}}_t = \widetilde{\mathbf{F}}_t = \mathbf{F}_t\mathbf{Q}_t$ gives

$$\operatorname{tr}\left[(\mathbf{F}_t\mathbf{Q}_t - \overline{\mathbf{F}}_t\mathbf{Q}_t)'\mathbf{W}^{-1}(\mathbf{F}_t\mathbf{Q}_t - \overline{\mathbf{F}}_t\mathbf{Q}_t)\right] \tag{A.111}$$

$$= \operatorname{tr}\left[\mathbf{Q}_t'(\mathbf{F}_t - \overline{\mathbf{F}}_t)'\mathbf{W}^{-1}(\mathbf{F}_t - \overline{\mathbf{F}}_t)\mathbf{Q}_t\right] \tag{A.112}$$

$$= \operatorname{tr}\left[\mathbf{Q}_t\mathbf{Q}_t'(\mathbf{F}_t - \overline{\mathbf{F}}_t)'\mathbf{W}^{-1}(\mathbf{F}_t - \overline{\mathbf{F}}_t)\right] \tag{A.113}$$

$$= \operatorname{tr}\left[(\mathbf{F}_t - \overline{\mathbf{F}}_t)'\mathbf{W}^{-1}(\mathbf{F}_t - \overline{\mathbf{F}}_t)\right] \tag{A.114}$$

where (A.113) and (A.114) follow from the cyclical property of the trace operator and the orthogonality of $\mathbf{Q}_t$. The expression in (A.114) matches the trace term in (A.104), which completes the proof. □

The following result will be useful in the proof of Proposition 2.

Proposition 7. *The Jacobian under the transformation f is given by*

$$J\left((\mathbf{A}_t, \mathbf{F}_t) \longrightarrow (\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t)\right) = 2^{-\frac{n(n+1)}{2}}\,|\det(\mathbf{H}_t)|^{\frac{m-1}{2}}. \tag{A.115}$$

Proof of Proposition 7. Letting $\mathbf{\Sigma}_t = \mathbf{H}_t^{-1}$, Arías et al. (2018) show that

$$J\left((\mathbf{A}_t, \mathbf{F}_t) \longrightarrow (\mathbf{B}_t, \mathbf{\Sigma}_t, \mathbf{Q}_t)\right) = 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{\Sigma}_t)|^{-\frac{2n+m+1}{2}} \tag{A.116}$$

By a known result, for example see Theorem 2.1.8 in Muirhead (1982), $J(\mathbf{\Sigma}_t \longrightarrow \mathbf{H}_t^{-1}) = \det(\mathbf{H}_t)^{-(n+1)}$. Hence,

$$J\left((\mathbf{A}_t, \mathbf{F}_t) \longrightarrow (\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t)\right) = \overbrace{2^{-\frac{n(n+1)}{2}} |\det(\mathbf{H}_t^{-1})|^{-\frac{2n+m+1}{2}}}^{J\left((\mathbf{A}_t,\mathbf{F}_t)\longrightarrow(\mathbf{B}_t,\mathbf{\Sigma}_t,\mathbf{Q}_t)\right)} \overbrace{|\det(\mathbf{H}_t)|^{-(n+1)}}^{J(\mathbf{\Sigma}_t \longrightarrow \mathbf{H}_t^{-1})} \tag{A.117}$$

$$= 2^{-\frac{n(n+1)}{2}} |\det(\mathbf{H}_t)|^{\frac{m-1}{2}} \tag{A.118}$$

□

Proof of Proposition 2.

Proof of Part I.

Given the density of $p\left(\mathbf{A}_t, \mathbf{F}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}\right)$ as derived in the proof of Theorem 1, I derive the distribution of $(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t) = f(\mathbf{A}_t, \mathbf{F}_t)$ as a change of variables. Recall that the density in the $(\mathbf{A}_t, \mathbf{F}_t)$-space is

$$\begin{aligned} p(\mathbf{A}_t, \mathbf{F}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}) &= p_{\mathbf{\Omega}}(\beta^{1/2} \mathbf{A}_{t-1}^{-1} \mathbf{A}_t) \cdot \beta^{n^2/2} |\mathbf{A}_{t-1}|^{-n} \\ & c \cdot \exp\{-\frac{1}{2} \mathrm{tr}[(\mathbf{F}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t)' \mathbf{W}^{-1} (\mathbf{F}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t)]\} \end{aligned} \tag{A.119}$$

where $c = (2\pi)^{-nm/2} |\mathbf{W}|^{-n/2}$. Next, I substitute in terms of $f(\mathbf{A}_t, \mathbf{F}_t)$. First, note that

$$\mathbf{F}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t = \mathbf{B}_t \mathbf{H}_t^{1/2} \mathbf{Q}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{H}_t^{1/2}\mathbf{Q}_t \tag{A.120}$$

$$= \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)\mathbf{H}_t^{1/2}\mathbf{Q}_t \tag{A.121}$$

implies

$$\mathrm{tr}\left[(\mathbf{F}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t)' \mathbf{W}^{-1} (\mathbf{F}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t)\right] \tag{A.122}$$

$$= \mathrm{tr}\left[\mathbf{Q}_t' \mathbf{H}_t^{1/2'} \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)' \mathbf{W}^{-1} \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right) \mathbf{H}_t^{1/2} \mathbf{Q}_t\right] \tag{A.123}$$

$$= \mathrm{tr}\left[\mathbf{H}_t^{1/2} \mathbf{Q}_t \mathbf{Q}_t' \mathbf{H}_t^{1/2'} \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)' \mathbf{W}^{-1} \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)\right] \tag{A.124}$$

$$= \mathrm{tr}\left[\mathbf{H}_t \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)' \mathbf{W}^{-1} \left(\mathbf{B}_t - \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)\right] \tag{A.125}$$

where the last equality follows from the cyclicality of the trace operator. Next, note that by Condition 1

$$p_{\boldsymbol{\Omega}}(\beta^{1/2}\mathbf{A}_{t-1}^{-1}\mathbf{H}_t^{1/2}\mathbf{Q}_t) = p_{\boldsymbol{\Omega}}\left(\beta^{1/2}\mathbf{A}_{t-1}^{-1}\mathbf{H}_t^{1/2}\right). \tag{A.126}$$

Putting these together, including the Jacobian gives

$$\begin{aligned} p(\mathbf{B}_t,\mathbf{H}_t,\mathbf{Q}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) &= p_{\boldsymbol{\Omega}}\left(\beta^{1/2}\mathbf{A}_{t-1}^{-1}\mathbf{H}_t^{1/2}\right)\cdot\beta^{n^2/2}|\mathbf{A}_{t-1}|^{-n} \\ &\cdot c\cdot\exp\{-\frac{1}{2}\mathrm{tr}\left[\mathbf{H}_t\left(\mathbf{B}_t-\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)'\mathbf{W}^{-1}\left(\mathbf{B}_t-\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\right)\right]\} \\ &\cdot J\left((\mathbf{A}_t,\mathbf{F}_t)\rightarrow(\mathbf{B}_t,\mathbf{H}_t,\mathbf{Q}_t)\right) \end{aligned} \tag{A.127}$$

From Proposition 7, the Jacobian term is not a function of $\mathbf{Q}_t$ and one can see from equation (A.127) that neither is any other term interacting with $(\mathbf{B}_t,\mathbf{H}_t)$, which proves that

$$\begin{aligned} &p(\mathbf{B}_t,\mathbf{H}_t,\mathbf{Q}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) \\ &\qquad= p(\mathbf{Q}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1})\cdot p(\mathbf{B}_t,\mathbf{H}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) \end{aligned} \tag{A.128}$$

Furthermore, no term explicitly involves $\mathbf{Q}_t$ and hence $p(\mathbf{Q}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) = p(\mathbf{Q}_t) = U(\mathcal{O}_n)$ and one can write

$$p(\mathbf{B}_t,\mathbf{H}_t,\mathbf{Q}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) = p(\mathbf{Q}_t)\cdot p(\mathbf{B}_t,\mathbf{H}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1}) \tag{A.129}$$

completing the proof of Part I.

Proof of Part II.

Shifting the time subscripts in equation (A.119) forward one period and substituting for $(\mathbf{A}_t,\mathbf{F}_t)$ as $\mathbf{B}_t,\mathbf{H}_t,\mathbf{Q}_t$ gives

$$\begin{aligned} p(\mathbf{A}_{t+1},\mathbf{F}_{t+1}|\boldsymbol{\phi},\mathbf{B}_t,\mathbf{H}_t,\mathbf{Q}_t) &= p_{\boldsymbol{\Omega}}(\beta^{1/2}\mathbf{Q}_t^{-1}\mathbf{H}_t^{-1/2}\mathbf{A}_{t+1})\cdot\beta^{n^2/2}|\mathbf{H}_t^{1/2}\mathbf{Q}_t|^{-n} \\ &c\cdot\exp\{-\frac{1}{2}\mathrm{tr}[(\mathbf{F}_{t+1}-\mathbf{B}_t\mathbf{A}_{t+1})'\mathbf{W}^{-1}(\mathbf{F}_{t+1}-\mathbf{B}_t\mathbf{A}_{t+1})]\} \end{aligned} \tag{A.130}$$

Note that

$$p_{\mathbf{\Omega}}(\beta^{1/2}\mathbf{Q}_t^{-1}\mathbf{H}_t^{-1/2}\mathbf{A}_{t+1}) = p_{\mathbf{\Omega}}(\mathbf{Q}_t^{-1}\beta^{1/2}\mathbf{H}_t^{-1/2}\mathbf{A}_{t+1}) \tag{A.131}$$

$$= p_{\mathbf{\Omega}}(\beta^{1/2}\mathbf{H}_t^{-1/2}\mathbf{A}_{t+1}) \tag{A.132}$$

where the last equality follows from the orthogonality of $\mathbf{Q}_t^{-1}$ and Condition 1. Noting also that $|\mathbf{H}_t^{1/2}\mathbf{Q}_t| = |\mathbf{H}_t^{1/2}||\mathbf{Q}_t|$ and $|\mathbf{Q}_t| = 1$, means that the density in equation (A.130) can be written without reference to $\mathbf{Q}_t$ as

$$\begin{aligned} p(\mathbf{A}_{t+1}, \mathbf{F}_{t+1}|\boldsymbol{\phi}, \mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t) &= p_{\mathbf{\Omega}}(\beta^{1/2}\mathbf{H}_t^{-1/2}\mathbf{A}_{t+1}) \cdot \beta^{n^2/2}|\mathbf{H}_t^{1/2}|^{-n} \\ & c \cdot \exp\{-\frac{1}{2}\mathrm{tr}[(\mathbf{F}_{t+1} - \mathbf{B}_t\mathbf{A}_{t+1})'\mathbf{W}^{-1}(\mathbf{F}_{t+1} - \mathbf{B}_t\mathbf{A}_{t+1})]\}\,. \end{aligned} \tag{A.133}$$

This completes the proof of Part II. □

Proof of Proposition 3. Substituting for $\mathbf{A}_{t-1}, \mathbf{F}_{t-1}$ in equation (A.119) gives

$$\begin{aligned} & p(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1}) \\ &\quad = p_{\mathbf{\Omega}}\big(\beta^{1/2}\mathbf{Q}_{t-1}^{-1}\mathbf{H}_{t-1}^{-1/2}\mathbf{H}_t^{1/2}\mathbf{Q}_t\big) \cdot \beta^{n^2/2}|\mathbf{H}_{t-1}^{1/2}\mathbf{Q}_{t-1}|^{-n} \\ &\quad\quad \cdot c \cdot \exp\{-\frac{1}{2}\mathrm{tr}\big[\mathbf{H}_t\big(\mathbf{B}_t - \mathbf{B}_{t-1}\big)'\mathbf{W}^{-1}\big(\mathbf{B}_t - \mathbf{B}_{t-1}\big)\big]\} \\ &\quad\quad \cdot 2^{-\frac{n(n+1)}{2}}|\det(\mathbf{H}_t)|^{\frac{m-1}{2}} \end{aligned} \tag{A.134}$$

The first line of which can be written as

$$p_{\mathbf{\Omega}}\big(\beta^{1/2}\mathbf{H}_{t-1}^{-1/2}\mathbf{H}_t^{1/2}\big) \cdot \beta^{n^2/2}|\mathbf{H}_{t-1}^{1/2}|^{-n} \tag{A.135}$$

and the final Jacobian term can be split as

$$|\det(\mathbf{H}_t)|^{\frac{m-1}{2}} = |\det(\mathbf{H}_t)|^{\frac{m}{2}}|\det(\mathbf{H}_t)|^{-\frac{1}{2}} \tag{A.136}$$

$$= |\det(\mathbf{H}_t^{-1})|^{-\frac{m}{2}}|\det(\mathbf{H}_t)|^{-\frac{1}{2}} \tag{A.137}$$

Noting that

$$\begin{aligned} & N(\mathbf{B}_t|\mathbf{B}_{t-1}, \mathbf{W}, \mathbf{H}_t^{-1}) \\ &\quad = |\det(\mathbf{H}_t^{-1})|^{-\frac{m}{2}} \cdot c \cdot \exp\{-\frac{1}{2}\mathrm{tr}\big[\mathbf{H}_t\big(\mathbf{B}_t - \mathbf{B}_{t-1}\big)'\mathbf{W}^{-1}\big(\mathbf{B}_t - \mathbf{B}_{t-1}\big)\big]\} \end{aligned} \tag{A.138}$$

Collecting these results, one can write the density as

$$
\begin{aligned}
&p(\mathbf{B}_t, \mathbf{H}_t, \mathbf{Q}_t | \boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_{t-1}, \mathbf{Q}_{t-1}) \\
&\quad = 2^{-\frac{n(n+1)}{2}} \cdot p_{\boldsymbol{\Omega}}\left(\beta^{1/2} \underline{h}(\mathbf{H}_{t-1})^{-1} \underline{h}(\mathbf{H}_t)\right) \cdot \beta^{n^2/2} |\underline{h}(\mathbf{H}_{t-1})|^{-n} |\mathbf{H}_t|^{-1/2} \\
&\quad\quad \cdot N(\mathbf{B}_t | \mathbf{B}_{t-1}, \mathbf{W}, \mathbf{H}_t^{-1})
\end{aligned}
\tag{A.139}
$$

Noting that the expression of the density makes no recourse to either $\mathbf{Q}_t$ or to $\mathbf{Q}_{t-1}$ proves the result. □

Proof of Proposition 4. The proof essentially amounts to a careful examination of equation (A.138). First, from the lack of dependence of any terms on $\mathbf{Q}_t$, it is apparent that $p(\mathbf{Q}_t) = U(\mathcal{O}_n)$. Second, the term $N(\mathbf{B}_t | \mathbf{B}_{t-1}, \mathbf{W}, \mathbf{H}_t^{-1})$ integrates to 1 regardless of $\mathbf{H}_t^{-1}$, leaving behind only the marginal density of $\mathbf{H}_t$, which depends on only $\mathbf{H}_{t-1}$ and $\boldsymbol{\phi}$. □

A.3 Densities for Reduced-Form Model

Proof of Proposition 5. First deriving the density of $\mathbf{H}_t$ as a change of variables from $\boldsymbol{\Gamma}_t$ gives

$$
p(\mathbf{H}_t | \boldsymbol{\phi}, h(\mathbf{H}_{t-1})) = p_{\boldsymbol{\Gamma}}\left(\underline{h}(\mathbf{H}_{t-1})^{-1'} \beta \mathbf{H}_t \underline{h}(\mathbf{H}_{t-1})^{-1}\right) \cdot J(\boldsymbol{\Gamma}_t \longrightarrow \mathbf{H}_t) \,. \tag{A.140}
$$

The Jacobian term is given by

$$
\begin{aligned}
J(\boldsymbol{\Gamma}_t \longrightarrow \mathbf{H}_t) &= |\beta^{1/2} \underline{h}(\mathbf{H}_{t-1})^{-1}|^{n+1} && \text{(A.141)} \\
&= (\beta^{n/2} |\underline{h}(\mathbf{H}_{t-1})^{-1}|)^{n+1} && \text{(A.142)} \\
&= \beta^{n(n+1)/2} |\underline{h}(\mathbf{H}_{t-1})|^{-(n+1)} \,. && \text{(A.143)}
\end{aligned}
$$

The expression for the Jacobian, which uses the fact that $\boldsymbol{\Gamma}_t$ is symmetric, is known in the literature and available from a variety of sources, such as Theorem 2.1.6. in Muirhead (1982). From here the result follows from applying Proposition 10 and simplifying. □

A.4 Properties of Shocks

Definition 3. *An $n \times n$ random matrix $\mathbf{Z}$ is distributed $B_n(\nu_1/2, \nu_2/2)$ for $\nu_1, \nu_2 > n-1$ if its density function is*

$$p(\mathbf{Z}) = c_{B_n(\nu_1/2,\nu_2/2)} \cdot \det(\mathbf{Z})^{(\nu_1-n-1)/2} \det(\mathbf{I}_n - \mathbf{Z})^{(\nu_2-n-1)/2}\,, \tag{A.144}$$

for $0 < \mathbf{Z} < \mathbf{I}_n$ and

$$c_{B_n(\nu_1/2,\nu_2/2)} = \frac{\Gamma_n\big((\nu_1+\nu_2)/2\big)}{\Gamma_n(\nu_1/2)\,\Gamma_n(\nu_2/2)} \tag{A.145}$$

Definition 4. *A random matrix $\mathbf{\Omega}$ is $HB_n(\nu_1/2, \nu_2/2)$-distributed, if it can be written as $\mathbf{\Omega} = \mathbf{Z}^{1/2}\mathbf{Q}$, where $\mathbf{Z} \sim B_n(\nu_1/2, \nu_2/2)$ and $\mathbf{Q} \sim U(\mathcal{O}_n)$ are independent.*

Proposition 8. *The density of $\mathbf{R} \sim HB_n(\nu_1/2, \nu_2/2)$, for $\nu_1, \nu_2 > n-1$, is*

$$p(\mathbf{R}) = c_{HB_n(\nu_1/2,\nu_2/2)} \cdot \det(\mathbf{R}\mathbf{R}')^{(\nu_1-n)/2} \det(\mathbf{I}_n - \mathbf{R}\mathbf{R}')^{(\nu_2-n-1)/2} \tag{A.146}$$

for $0 < \mathbf{R}\mathbf{R}' < \mathbf{I}_n$ and

$$c_{HB_n(\nu_1/2,\nu_2/2)} = \frac{\Gamma_n(n/2)}{\pi^{n^2/2}} \cdot \underbrace{\left[\frac{\Gamma_n\big((\nu_1+\nu_2)/2\big)}{\Gamma_n(\nu_1/2)\,\Gamma_n(\nu_2/2)}\right]}_{c_{B_n(\nu_1/2,\nu_2/2)}} \tag{A.147}$$

Proof of Proposition 8. The proof proceeds along similar lines as Theorem 5.3.21 in Gupta and Nagar (2000). Starting from the joint density of the two random matrices invoked in Definition 4, the joint density of $\mathbf{Z}$ and $\mathbf{Q}$ is

$$p(\mathbf{Z}, \mathbf{Q}) = c_{\mathbf{Z}} c_{\mathbf{Q}} \cdot \det(\mathbf{Z})^{(\nu_1-n-1)/2} \det(\mathbf{I}_n - \mathbf{Z})^{(\nu_2-n-1)/2} \cdot g_{n,n}(\mathbf{Q})\,, \tag{A.148}$$

where $c_{\mathbf{Z}} = c_{B_n(\nu_1/2,\nu_2/2)}$ and $c_{\mathbf{Q}} = \Gamma_n(n/2)/(2^n\pi^{n^2/2})$. Transform $\mathbf{Z} = \mathbf{T}\mathbf{T}'$ for $\mathbf{T}$

lower triangular with positive diagonal elements, with the Jacobian given by

$$J(\mathbf{Z} \longrightarrow \mathbf{T}\mathbf{T}') = 2^n \prod_{t=1}^{n} t_{ii}^{n-i+1} \tag{A.149}$$

and where t_{ii} denotes the element of $\mathbf{T}$ in row-i, column-i.

Next, consider the transformation $\mathbf{R} = \mathbf{T}\mathbf{Q}$, with the Jacobian given by

$$J(\mathbf{T}, \mathbf{Q} \longrightarrow \mathbf{R}) = \prod_{i=1}^{n} t_{ii}^{-n+i} \left[g_{n,n}(\mathbf{Q})\right]^{-1} \tag{A.150}$$

Hence, the Jacobian of the transformation

$$J(\mathbf{Z}, \mathbf{Q} \longrightarrow \mathbf{R}) = J(\mathbf{Z} \longrightarrow \mathbf{T}\mathbf{T}')\, J(\mathbf{T}, \mathbf{Q} \longrightarrow \mathbf{R}) \tag{A.151}$$

$$= \left[2^n \prod_{t=1}^{n} t_{ii}^{n-i+1}\right] \left[\prod_{i=1}^{n} t_{ii}^{-n+i} \left[g_{n,n}(\mathbf{Q})\right]^{-1}\right] \tag{A.152}$$

$$= 2^n \left(\prod_{i=1}^{n} t_{ii}\right) \left[g_{n,n}(\mathbf{Q})\right]^{-1} \tag{A.153}$$

Noting that $|\det(\mathbf{R})| = \prod_{i=1}^{n} t_{ii}$ gives

$$J(\mathbf{Z}, \mathbf{Q} \longrightarrow \mathbf{R}) = 2^n |\det(\mathbf{R})| \left[g_{n,n}(\mathbf{Q})\right]^{-1} \tag{A.154}$$

$$= 2^n \det(\mathbf{R}\mathbf{R}')^{1/2} \left[g_{n,n}(\mathbf{Q})\right]^{-1} \tag{A.155}$$

The density of $\mathbf{R}$ is then

$$p(\mathbf{R}) = c_{\mathbf{Z}} c_{\mathbf{Q}} \cdot \det(\mathbf{R}\mathbf{R}')^{(\nu_1-n-1)/2} \det(\mathbf{I}_n - \mathbf{R}\mathbf{R}')^{(\nu_2-n-1)/2} \cdot 2^n \det(\mathbf{R}\mathbf{R}')^{1/2}, \tag{A.156}$$

$$= c_{\mathbf{R}} \cdot \det(\mathbf{R}\mathbf{R}')^{(\nu_1-n)/2} \cdot \det(\mathbf{I}_n - \mathbf{R}\mathbf{R}')^{(\nu_2-n-1)/2} \tag{A.157}$$

where $c_{\mathbf{R}} = 2^n c_{\mathbf{Z}} c_{\mathbf{Q}}$, which simplifies to the expression given in Proposition 8.[23]

[23] Note that in the special case of $\nu_1 = n$, the distribution of the random matrix becomes that of an "inverted matrix variate t-distribution" with parameters $IT_{n,n}(\nu_2 - n + 1, \mathbf{0}_{n,n}, \mathbf{I}_n, \mathbf{I}_n)$, where the notation is that of GN's Definition 4.4.1. In that case the proof is the same as that of GN's Theorem 5.3.21.

□

The distribution of $\mathbf{\Omega}$ satisfies the key condition to admit a reduced form.

Proposition 9. *The distribution of a random matrix* $\mathbf{\Omega} \sim HB(\nu_1/2, \nu_2/2)$ *satisfies Condition 1.*

Proof of Proposition 9. The proof follows by evaluating the density in Proposition 8 directly at the point in $\mathbf{Q\Omega P}$ for $\mathbf{Q}, \mathbf{P} \in \mathcal{O}_n$.

$$p(\mathbf{Q\Omega P}) = c_{HB_n(\nu_1/2,\nu_2/2)} \cdot \det(\mathbf{Q\Omega PP'\Omega'Q'})^{(\nu_1-n)/2} \cdot \det(\mathbf{I}_n - \mathbf{Q\Omega PP'\Omega'Q'})^{(\nu_2-n-1)/2}. \tag{A.158}$$

Noting that the determinant terms simplify as $\det(\mathbf{Q\Omega PP'\Omega'Q'}) = \det(\mathbf{Q\Omega\Omega'Q'}) = \det(\mathbf{Q})\det(\mathbf{\Omega\Omega'})\det(\mathbf{Q'}) = \det(\mathbf{\Omega\Omega'})$ and $\det(\mathbf{I}_n - \mathbf{Q\Omega PP'\Omega'Q'}) = \det\left(\mathbf{Q}(\mathbf{I}_n - \mathbf{\Omega PP'\Omega'})\mathbf{Q'}\right) = \det\left(\mathbf{Q}(\mathbf{I}_n - \mathbf{\Omega\Omega'})\mathbf{Q'}\right) = \det(\mathbf{Q})\det(\mathbf{I}_n - \mathbf{\Omega\Omega'})\det(\mathbf{Q'}) = \det(\mathbf{I}_n - \mathbf{\Omega\Omega'})$ gives

$$p(\mathbf{Q\Omega P}) = c_{HB_n(\nu_1/2,\nu_2/2)} \cdot \det(\mathbf{\Omega\Omega'})^{(\nu_1-n)/2} \det(\mathbf{I}_n - \mathbf{\Omega\Omega'})^{(\nu_2-n-1)/2} \tag{A.159}$$

$$= p(\mathbf{\Omega}). \tag{A.160}$$

□

Proposition 10. *If the distribution of* $\mathbf{\Omega}$ *satisfies Condition 1, with pdf denoted* $p_{\mathbf{\Omega}}$*, then the density of* $\mathbf{\Gamma} = \mathbf{\Omega\Omega}$ *is*

$$p(\mathbf{\Gamma}) = \frac{\pi^{n^2/2}}{\Gamma_n(n/2)} |\det(\mathbf{\Gamma})|^{-1/2} p_{\mathbf{\Omega}}\left(\mathbf{T}_{\mathbf{\Gamma}}\right), \tag{A.161}$$

where $\mathbf{T}_{\mathbf{\Gamma}}$ *denotes the unique lower triangular* $n \times n$ *factorization, with positive diagonal elements, of* $\mathbf{\Gamma}$ *such that* $\mathbf{T}_{\mathbf{\Gamma}}\mathbf{T}_{\mathbf{\Gamma}}' = \mathbf{\Gamma}$.

Proof of Proposition 10. The proof follows a standard line of argument; see, for example, Theorem 3.2.2 in Gupta and Nagar (2000). First, make the transformation $\mathbf{\Omega} = \mathbf{TL}$ for $\mathbf{T}$ lower triangular with positive diagonal and $\mathbf{L} \in \mathcal{O}_n$. The Jacobian is given by $J(\mathbf{\Omega} \to \mathbf{T}, \mathbf{L}) = \prod_{i=1}^{n} t_{ii}^{n-i} g_{n,n}(\mathbf{L})$, which gives the joint density of $(\mathbf{T}, \mathbf{L})$ as

$$p(\mathbf{T}, \mathbf{L}) = p_{\mathbf{\Omega}}(\mathbf{TL}) \prod_{i=1}^{n} t_{ii}^{n-i} g_{n,n}(\mathbf{L}). \tag{A.162}$$

By Condition 1 $p_{\mathbf{\Omega}}(\mathbf{TL}) = p_{\mathbf{\Omega}}(\mathbf{T})$, so one can write

$$p(\mathbf{T}, \mathbf{L}) = p_{\mathbf{\Omega}}(\mathbf{T}) \prod_{i=1}^{n} t_{ii}^{n-i} g_{n,n}(\mathbf{L}). \tag{A.163}$$

Next integrate out $\mathbf{L}$ to obtain the marginal density of $\mathbf{T}$ as

$$p(\mathbf{T}) = p_{\mathbf{\Omega}}(\mathbf{T}) \prod_{i=1}^{n} t_{ii}^{n-i} \cdot \underbrace{\int_{\mathbf{LL}'=\mathbf{I}_n} g_{n,n}(\mathbf{L})\, d\mathbf{L}}_{\frac{2^n \pi^{(1/2)n^2}}{\Gamma_n(n/2)}}. \tag{A.164}$$

Lastly, transform $\mathbf{\Gamma} = \mathbf{TT}'(= \mathbf{\Omega\Omega}')$, where $J(\mathbf{T} \to \mathbf{\Gamma}) = \left(2^n \prod_{i=1}^{n} t_{ii}^{n-i+1}\right)^{-1}$. After some simplifying and noting that $|\det(\mathbf{\Gamma})|^{-1/2} = \prod_{i=1}^{n} t_{ii}^{-1}$, the density of $\mathbf{\Gamma}$ is

$$p(\mathbf{\Gamma}) = \frac{\pi^{n^2/2}}{\Gamma_n(n/2)} |\det(\mathbf{\Gamma})|^{-1/2} p_{\mathbf{\Omega}}(\mathbf{T}_{\mathbf{\Gamma}}). \tag{A.165}$$

□

From Proposition 10 one can easily obtain the density for the reduced-form law of motion induced by "half matrix beta" structural shocks.

Proposition 11. *If* $\mathbf{\Omega} \sim HB_n(\nu_1/2, \nu_2/2)$*, then* $\mathbf{\Gamma} = \mathbf{\Omega\Omega}'$ *is such that* $\mathbf{\Gamma} \sim B_n(\nu_1/2, \nu_2/2)$.

Proof of Proposition 11. Noting that $\mathbf{T}_{\mathbf{\Gamma}}\mathbf{T}_{\mathbf{\Gamma}}' = \mathbf{\Omega\Omega}' = \mathbf{\Gamma}$ and applying Proposition 10 to the density in Definition 4, the density of $\mathbf{\Gamma}$ is given as

$$p(\mathbf{\Omega}) = \frac{\pi^{n^2/2}}{\Gamma_n(n/2)} \det(\mathbf{\Gamma})^{-1/2} \left[c \cdot \frac{\Gamma_n(n/2)}{\pi^{n^2/2}} \det(\mathbf{\Gamma})^{\frac{\nu_1-n}{2}} \det(\mathbf{I}_n - \mathbf{\Gamma})^{\frac{\nu_2-n-1}{2}} \right] \tag{A.166}$$

where $c = c_{B_n(\nu_1/2, \nu_2/2)}$. Simplifying yields the density in Definition 3. □

B. Gibbs Sampler for DLM-DWSV Parameters

This section of the appendix gives the details of the steps of the Gibbs sampler.

Initialization. I initialize the MCMC algorithm by simulating many random draws from the prior for the static parameters $\boldsymbol{\phi}$ and evaluating their marginal posterior kernel. I then choose the value for $\boldsymbol{\phi}$ with the highest value for the marginal posterior kernel, call it $\boldsymbol{\phi}^*$. I then simulate a sequence of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ backwards conditional on $\boldsymbol{\phi}^*$ as described in Table II.

Block 1: $\mathbf{W} \mid \mathbf{y}_{1:T}, \beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}$

Given a draw of the history of latent states $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$, each matrix of shocks $\mathbf{V}_t$ to the linear coefficients is known via

$$\mathbf{V}_t = \mathbf{B}_t - \mathbf{B}_{t-1}. \tag{B.167}$$

Assuming the prior has the form $p(\mathbf{W}|\beta) \sim IW(\Psi_0, \nu_0)$, the conditional posterior of $\mathbf{W}$ is

$$p(\mathbf{W}|Y, \beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T}) \sim IW(\Psi_{0:T}, \nu_{0:T}) \tag{B.168}$$

where

$$\Psi_{0:T} = \Psi_0 + \Psi_{1:T} \tag{B.169}$$

$$\nu_{0:T} = \nu_0 + \nu_{1:T} \tag{B.170}$$

and

$$\Psi_{1:T} = \sum_{t=1}^{T} \mathbf{V}_t \mathbf{H}_t \mathbf{V}_t' \tag{B.171}$$

$$\nu_{1:T} = \sum_{t=1}^{T} n = Tn\,. \tag{B.172}$$

Block 2: $\beta, \mathbf{B}_{0:T}, \mathbf{H}_{0:T} \mid \mathbf{y}_{1:T}, \mathbf{W}$

Sampling from Block 2 entails sampling from the joint distribution of β, $\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\mathbf{y}_{1:T}, \mathbf{W}$. I accomplish this by first sampling from the distribution of

$\beta|\mathbf{y}_{1:T}, \mathbf{W}$, which is marginal of $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$, and subsequently conditioning on β and sampling from the distribution of $\mathbf{B}_{0:T}, \mathbf{H}_{0:T}|\mathbf{y}_{1:T}, \mathbf{W}, \beta$.

Step 2a: $\beta \mid \mathbf{y}_{1:T}, \mathbf{W}$

The form of this step is often referred to as a Metropolis-within-Gibbs step. Given $\beta^{(i-1)}$, one "proposes" a value for $\beta^{(i)}$, call the proposal β^*, as a random sample from a density $q(\beta^*|\beta^{(i-1)})$. One "accepts" β^* and sets $\beta^{(i)} = \beta^*$ with probability

$$\alpha\left(\beta^*|\mathbf{y}_{1:T}, \mathbf{W}\right) = \min\left\{\frac{p\left(\beta^*, \mathbf{W}^{(i)}|\mathbf{y}_{1:T}\right) q(\beta^{(i-1)}|\beta^*)}{p\left(\beta^{(i-1)}, \mathbf{W}^{(i)}|\mathbf{y}_{1:T}\right) q(\beta^*|\beta^{(i-1)})},\ 1\right\}. \tag{B.173}$$

If β^* is rejected, one sets $\beta^{(i)} = \beta^{(i-1)}$. I use $q(\beta^*|\beta^{(i-1)}) = p_N(\beta^{(i-1)}, \sigma_\beta)$, which is symmetric and hence the ratio of $q(\cdot)$ densities in (B.173) cancels. Let

$$k\left(\mathbf{W}, \beta|\mathbf{y}_{1:T}\right) = p(\mathbf{W}, \beta)\, p(\mathbf{y}_{1:T}|\beta, \mathbf{W}), \tag{B.174}$$

where $k(\mathbf{W}, \beta|\mathbf{y}_{1:T})$ differs from $p(\mathbf{W}, \beta|\mathbf{y}_{1:T})$ by only a normalizing constant that would cancel in (B.173). Hence, we can calculate α as

$$\alpha\left(\beta^*|\mathbf{y}_{1:T}, \mathbf{W}\right) = \min\left\{\frac{k\left(\mathbf{W}^{(i)}, \beta^*|\mathbf{y}_{1:T}\right)}{k\left(\mathbf{W}^{(i)}, \beta^{(i-1)}|\mathbf{y}_{1:T}\right)},\ 1\right\}, \tag{B.175}$$

so long as we can calculate $k(\mathbf{W}, \beta|\mathbf{y}_{1:T})$ pointwise. To evaluate the kernel, I presume one can evaluate $p(\mathbf{W}, \beta)$ pointwise. One can evaluate $p(\mathbf{y}_{1:T}|\beta, \mathbf{W})$ pointwise by using the recursive filtering algorithm summarized in Table I computing

$$p(\mathbf{y}_{1:T}|\beta, \mathbf{W}) = \prod_{t=1}^{T} \underbrace{p(\mathbf{y}_t|\mathbf{W}, \beta, \mathbf{y}_{1:t-1})}_{\text{Step 1.5 in Table I}} \tag{B.176}$$

Step 2b: $\mathbf{B}_{0:T}, \mathbf{H}_{0:T} \mid \mathbf{y}_{1:T}, \beta, \mathbf{W}$

The final step is a sample from the recursive backwards simulation "smoother" algorithm for $(\mathbf{B}_{0:T}, \mathbf{H}_{0:T})$ summarized in Table II. The draw proceeds backwards from the end of the forward filtering algorithm.

References

Arías, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2018). "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications." *Econometrica*, 86(2), pp. 685–720. doi:10.3982/ECTA14468.

Baumeister, Christiane and Luca Benati (2013). "Unconventional monetary policy and the great recession: estimating the macroeconomic effects of a spread compression at the zero lower bound." *International Journal of Central Banking*, 9(2), pp. 165–211.

Baumeister, Christiane and James D. Hamilton (2015). "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83(5), pp. 1963–1999. doi:10.3982/ECTA12356.

Baumeister, Christiane and Gert Peersman (2013a). "The role of time-varying price elasticities in accounting for volatility changes in the crude oil market." *Journal of Applied Econometrics*, 28(7), pp. 1087–1109. doi:10.1002/jae.2283.

Baumeister, Christiane and Gert Peersman (2013b). "Time-varying effects of oil supply shocks on the US economy." *American Economic Journal: Macroeconomics*, 5(4), pp. 1–28. doi:10.1257/mac.5.4.1.

Canova, Fabio and Gianni De Nicolo (2002). "Monetary disturbances matter for business fluctuations in the G-7." *Journal of Monetary Economics*, 49(6), pp. 1131–1159. doi:10.1016/S0304-3932(02)00145-9.

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(2), pp. 477–490. doi:10.1016/j.jedc.2008.05.010.

Clark, Todd E. and Francesco Ravazzolo (2015). "Macroeconomic forecasting performance under alternative specifications of time-varying volatility." *Journal of Applied Econometrics*, 30(4), pp. 551–575. doi:10.1002/jae.2379. URL `http://dx.doi.org/10.1002/jae.2379`.

Cogley, Timothy and Thomas J. Sargent (2001). "Evolving post-World War II U.S. inflation dynamics." *NBER Macroeconomics Annual*, 16, pp. 331–373. doi:10.1086/654451.

Cogley, Timothy and Thomas J. Sargent (2005). "Drifts and volatilities: monetary policies and outcomes in the post WWII US." *Review of Economic Dynamics*, 8(2), pp. 262 – 302. doi:10.1016/j.red.2004.10.009.

Del Negro, Marco and Giorgio E. Primiceri (2015). "Time varying structural vector autoregressions and monetary policy: A corrigendum." *The Review of Economic Studies*, 82(4), pp. 1342–1345. doi:10.1093/restud/rdv024.

Del Negro, Marco and Frank Schorfheide (2011). "Bayesian macroeconometrics." In John Geweke, Gary Koop, and Herman van Dijk, editors, *The Oxford Handbook of Bayesian Econometrics*, chapter 7, pp. 293–389. Oxford University Press. doi:10.1093/oxfordhb/9780199559084.013.0008.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984). "Forecasting and conditional projection using realistic prior distributions." *Econometric Reviews*, 3(1), pp. 1–100. doi:10.1080/07474938408800053.

Faust, Jon (1998). "The robustness of identified VAR conclusions about money." *Carnegie-Rochester Conference Series on Public Policy*, 49, pp. 207 – 244. doi:https://doi.org/10.1016/S0167-2231(99)00009-3.

Fox, E. B. and M. West (2014). "Autoregressive models for variance matrices: Stationary inverse Wishart processes." URL `http://ftp.stat.duke.edu/WorkingPapers/11-15.html`, arXiv:1107.5239.

Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*. Chapman & Hall/CRC.

Harrison, Jeff and Mike West (1997). *Bayesian Forecasting & Dynamic Models*. Springer.

Hofmann, Boris, Gert Peersman, and Roland Straub (2012). "Time variation in U.S. wage dynamics." *Journal of Monetary Economics*, 59(8), pp. 769–783. doi:10.1016/j.jmoneco.2012.10.009.

Konno, Yoshihiko (1988). "Exact moments of the multivariate f and beta distributions." *Journal of The Japan Statistical Society*, 18(2), pp. 123–130. doi:10.11329/jjss1970.18.123.

Koop, Gary and Dimitris Korobilis (2013). "Large time-varying parameter VARs." *Journal of Econometrics*, 177(2), pp. 185–198. doi:10.1016/j.jeconom.2013.04.007.

Mountford, Andrew and Harald Uhlig (2009). "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics*, 24(6), pp. 960–992. doi:10.1002/jae.1079. URL `10.1002/jae.1079`.

Muirhead, Robb J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc.

Prado, Raquel and Mike West (2010). *Time Series: Modeling, Computation, and Inference*. Chapman & Hall/CRC.

Primiceri, Giorgio E. (2005). "Time varying structural vector autoregressions and monetary policy." *The Review of Economic Studies*, 72(3), pp. 821–852. doi:10.1111/j.1467-937X.2005.00353.x.

Quintana, J., V. Lourdes, O. Aguilar, and J. Liu (2003). "Global gambling." In J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, editors, *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, pp. 349–368. Oxford University Press.

Quintana, José M and Mike West (1987). "An analysis of international exchange rates using multivariate DLM's." *The Statistician*, pp. 275–281. doi:10.2307/2348524.

Rothenberg, Thomas J. (1971). "Identification in parametric models." *Econometrica*, 39(3), pp. 577–591. doi:10.2307/1913267. URL `http://www.jstor.org/stable/1913267`.

Rubio-Ramírez, Juan F., Daniel F. Waggoner, and Tao Zha (2010). "Structural vector autoregressions: Theory of identification and algorithms for inference." *The Review of Economic Studies*, 77(2), pp. 665–696. doi:10.1111/j.1467-937X.2009.00578.x.

Sims, Christopher A (1993). "A nine-variable probabilistic macroeconomic forecasting model." In *Business Cycles, Indicators and Forecasting*, pp. 179–212. University of Chicago Press.

Sims, Christopher A., Daniel F. Waggoner, and Tao Zha (2008). "Methods for inference in large multiple-equation Markov-switching models." *Journal of Econometrics*, 146(2), pp. 255 – 274. doi:10.1016/j.jeconom.2008.08.023.

Sims, Christopher A. and Tao Zha (2006). "Were there regime switches in U.S. monetary policy?" *The American Economic Review*, 96(1), pp. 54–81. doi:10.1257/000282806776157678.

Stewart, Gilbert W (1980). "The efficient generation of random orthogonal matrices with an application to condition estimators." *SIAM Journal on Numerical Analysis*, 17(3), pp. 403–409.

Uhlig, Harald (1994). "On singular Wishart and singular multivariate beta distributions." *The Annals of Statistics*, 22(1), pp. 395–405. doi:10.1214/aos/1176325375.

Uhlig, Harald (1997). "Bayesian vector autoregressions with stochastic volatility." *Econometrica*, 65(1), pp. pp. 59–73. doi:10.2307/2171813.

Uhlig, Harald (2005). "What are the effects of monetary policy on output? Results from an agnostic identification procedure." *Journal of Monetary Economics*, 52(2), pp. 381 – 419. doi:10.1016/j.jmoneco.2004.05.007.