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

Mark Bognanni*
Federal Reserve Bank of Cleveland

This Draft: November 22, 2017, **Preliminary**

Abstract. I develop a new class of structural vector autoregressions with time-varying parameters. Unlike similar models in the literature, my model's law of motion for the time-varying parameters yields a reduced-form representation that is shared by all structural models in the class. Structural inference can then proceed by first estimating the reduced-form parameters and then mapping them to structural parameters by imposing exact—or partial—identification restrictions. I provide an algorithm for fully-Bayesian estimation of the reduced-form model and apply the framework to the estimation of time-varying fiscal multipliers under partial identification.

JEL: C11, C15, C32, C52, E3, E4, E5
Keywords: structural vector autoregressions, time-varying parameters, Gibbs sampling, stochastic volatility, Bayesian inference

---

I thank Christiane Baumeister, Dario Caldara, Domenico Giannone, Yuriy Gorodnichenko, Jim Hamilton, Eric Leeper, Karel Mertens, Juan Rubio-Ramírez, Frank Schorfheide, Dan Waggoner, and Mike West, as well as my Cleveland Fed colleagues Dionissi Aliprantis, Todd Clark, Ed Knotek, Ellis Tallman, and Saeed Zaman, for helpful comments and conversations. I thank John Zito for outstanding research assistance. The views expressed in this paper do not necessarily reflect those of Federal Reserve Bank of Cleveland, the Federal Reserve Board of Governors, or the Federal Reserve System.

*Mark Bognanni, Economic Research Department, Federal Reserve Bank of Cleveland, PO Box 6387, Cleveland, OH 44101-1387 (Email: `email.markbognanni@gmail.com`, Web: `http://markbognanni.com`).

1. Introduction

The last two decades of research on structural vector autoregressions (SVARs) has largely pursued methods for 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 relaxing the traditional types of identifying restrictions has followed from the seminal contributions on partially identified SVARs of Canova and De Nicolo (2002) and Uhlig (2005). To date, however, these two research agendas have lived largely separate lives.

This paper's primary contribution is to be the first to merge structural time-varying-parameter dynamic systems and partial identification in an internally consistent probabilistic framework. My TVP-SVAR admits a reduced-form for which I provide a tractable MCMC algorithm for inference. Following estimation of the reduced-form parameters, the researcher may then apply the existing methods from the constant-parameter SVAR framework off-the-shelf, including partial identification methods, while preserving an internally consistent probabilistic model. Hence, one might alternatively summarize this paper's key contribution as follows: it provides a TVP-SVAR amenable to the wholesale extension of the widely-used methods developed in Rubio-Ramírez, Waggoner, and Zha (2010) to a TVP setting.

The key challenge in constructing such a model is in specifying laws of motion for the dynamic parameters that maintain the invariance of the likelihood function to orthogonal rotations of the dynamic parameters. This invariance property is the key element that justifies the widely-used two-step approach for the estimation of constant-parameter SVARs under either exact or partial

[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).

identification in which researchers first estimate the reduced-form parameters that would be implied by the parameters of any candidate structural system and then find orthogonal rotations of the reduced-form parameters that yields structural parameters satisfying the identifying restrictions.[2] This approach yields valid, likelihood-based inference for the SVAR because of the observational equivalence of the reduced-form and all candidate SVARs, the space of which is conveniently indexed by orthogonal matrices. In the case of exact identifying restrictions there is a unique value for valid structural parameters that are observationally equivalent to given reduced-form parameters, while in the case of sign-restrictions as in Uhlig (2005), a set of valid structural parameters are observationally equivalent to the reduced-form parameters.

This paper's contributions are necessary because the models developed in Cogley and Sargent (2005) and Primiceri (2005) do not satisfy the invariance property. Nonetheless, researchers have sought to use the partial-identification toolkit to infer time-varying objects of interest in a fashion analogous to that used in constant-parameter SVARs: estimating a VAR-TVP-SV, constructing the "reduced-form" VAR system period-by-period, draw-by-draw, and applying standard sign-restrictions methods to decompose the "reduced-form" VAR parameters, again period-by-period, draw-by-draw.[3] However the results based on arbitrary orthogonal rotations of VAR-TVP-SV's parameters are, at best, difficult to interpret since in the context of the probability model originally estimated, the data can in fact tell the difference between the alternative parameter values. One particularly salient way to grasp the lack of the invariance to orthogonal rotations is to note that the results from the VAR-TVP-SV model can, in principle, depend on the ordering of variables, even if one only cares the "reduced-form parameters" implied by the structural parameters.[4] While clearly conceptually

[2]When the restrictions are either exactly or partially identifying, then at least one such rotation is guaranteed to exist for almost any value of the reduced-form parameters.

[3]Such an approach to structural inference has been employed in Canova and Gambetti (2009), Baumeister and Peersman (2013b), Baumeister and Peersman (2013a), and Hofmann, Peersman, and Straub (2012).

[4]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 (2014) for a discussion of this issue. Note that both Cogley and Sargent (2005) and Primiceri

undesirable, the potential sensitivity to variable ordering in the VAR-TVP-SV is particularly problematic in practice because an n-variable VAR admits $n!$ unique orderings, and estimating even a single specification of the model is computationally demanding.[5]

In contrast, the DSVAR I evelop specifies laws of motion for the time-varying parameters whose likelihood function is invariant to orthogonal rotations of any elements in the *sequences* of structural parameters. To be precise, given estimates of the reduced-form parameters, all *sequences* of possible structural parameters that differ by an orthogonal rotation are, ex post, equi-probable. The model specifies laws of motion for stochastic time-variation in the reduced-form system directly and is known in the Bayesian statistics literature as the dynamic linear model with discounted-Wishart stochastic volatility (DLM-DWSV). Key for my purposes, the DLM-DWSV model's likelihood function is invariant to period-by-period orthogonal rotations of a matrix square root of the time-varying covariance matrix. One implication of this property is that the estimated system is invariant to variable reordering, since such permutations are merely special cases of orthogonal rotations. As a result of this property, one can, in principle, estimate the DLM-DWSV model and then apply identifying sign restrictions for identification in a post-processing stage.

This paper builds off of a number of papers more familiar to statisticians than to economists. My DSVAR yields a reduced-form 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),

---

(2005) work with small dynamic systems of three variables and two lags, which make it feasible to check robustness against all six possible orderings, though it remains unclear how one would consider partial identification.

[5]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. Since an n-variable VAR admits $n!$ unique orderings, a researcher using only a modest 4 variable system, like that in Baumeister and Peersman (2013b), confronts 24 possible orderings, a thorough exploration of which strains the current limits of computational feasibility. A researcher daring enough to consider a 5 variable system would be faced with 120 possible orderings, leaving the researcher no choice but to plead ignorance regarding the importance of variable ordering for their results.

while the discounted-Wishart stochastic volatility process was formalized as a valid probability model by Uhlig (1994) and Uhlig (1997). Prado and West (2010) gives the most thorough treatment to date of the complete model, including a description of the posterior smoothing algorithm for the dynamic parameters, which is critical for my estimation algorithm.

In addition to its primary contribution, this paper makes two additional contributions purely in the context of the SVAR's reduced-form. First, I contribute directly to the literature on DLMs by presenting an MCMC algorithm for fully-Bayesian inference for all model parameters, including the discounting parameter governing the variability of the stochastic volatility. Koop and Korobilis (2013) consider forecasting with a model similar to the reduced-form DLM-DWSV but without likelihood-based estimation of the model 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. Second, the paper demonstrates that DLM-DWSV can handle dynamic systems considerably larger than those previously considered in the literature estimating the VAR-TVP-SV; the appliation on time-varying fiscal multipliers uses $n = 7$ variables and $p = 4$ lags.

The methods I develop in this paper are fully complementary to a number of other recent contributions on SVAR identification, particularly Amir-Ahmadi and Drautzburg (2017) and Antolín-Díaz and Rubio-Ramírez (2017).

From here the rest of the paper proceeds as follows. In Section 2, I review the current frameworks for partial or exact identifiction in SVARs to facilitate comparison with my TVP extension. In Section 3, I describe the reduced-form model with time-varying parameters and stochastic volatility. In Section 4 I describe the MCMC algorithm for estimating the reduced-form model. In Section 5 I confront the model and estimation procedure with an empirical application based on estimating time-varying fiscal multipliers. In Section 6 I conclude.

Lastly before moving on I introduce a few notational conventions used throughout the paper. Matrices are styled as uppercase bold, as in $\mathbf{A}$. Vectors are styled as lowercase bold, as in $\mathbf{y}$. Coefficients are styled as lowercase normal weight, as in β.

2. A crash course in structural VARs (and the reduced-form VARs inside of which they hide)

To fix ideas, in this section I describe the standard constant-parameter SVAR framework. My notation and description of the framework closely follow that of Rubio-Ramírez et al. (2010), to which I refer readers for further details.

2.1 The Structural Model

I treat the structural VAR as the primitive. I assume that the structural model describing the evolution of n observable and endogenous economic variables linearly relates each variable to the contemporaneous and lagged values of all other 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

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

where $\boldsymbol{\varepsilon}_t$ is an $(n \times 1)$ vector of exogenous structural shocks, "N" denotes the Gaussian distribution, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. 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 the invertability of $\mathbf{A}$.

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

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

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 exogenous disturbances and I define $\mathbf{S} \equiv (\mathbf{A}, \mathbf{F})$ to refer to the tuple of structural parameters. In the absence of further restrictions, the space of possible values for $\mathbf{S}$ is the subset of $\mathbb{R}^{mn+n^2}$ for which $\mathbf{A}$ is invertible, and I refer to this space as $\mathcal{S}^U$ where the superscript "U" indicates that this is the "unrestricted" parameter space. Letting $p_{\mathbf{S}}(\cdot)$ refer to the data density for $\mathbf{y}_{1:T}$ implied by equation (2), I use the notation S^U to refer to the unrestricted structural model, i.e. $S^U = (p_{\mathbf{S}}(\cdot), \mathbf{S}, \mathcal{S}^U)$.

The objects of interest to the economist are typically either $\mathbf{S}$ or a function of $\mathbf{S}$, such as impulse responses or variance decompositions with respect to particular shocks in $\boldsymbol{\varepsilon}_t$. Unfortunately, without further restrictions on the parameter space, the data cannot identify the elements of $\mathbf{S}$: for any $(\mathbf{A}, \mathbf{F}) = \mathbf{S} \in \mathcal{S}^U$ and any $n \times n$ orthogonal matrix $\mathbf{Q}$, the alternative parameters defined as $\tilde{\mathbf{S}} = (\mathbf{AQ}, \mathbf{FQ})$ are observationally equivalent to $(\mathbf{A}, \mathbf{F})$.[6] To be precise, and echoing Rothenberg (1971), I characterize observational equivalence in terms of data densities as follows.

Definition 1 (Observational equivalence of data densities)**.** *If $p(\cdot)$ and $\tilde{p}(\cdot)$ are data densities for observables $\mathbf{y}_{1:T}$, then $p(\cdot)$ and $\tilde{p}(\cdot)$ are observationally equivalent if and only if $p(\mathbf{y}_{1:T}) = \tilde{p}(\mathbf{y}_{1:T})$ for any $\mathbf{y}_{1:T}$.*

One can, however, identify certain *combinations* of parameters in $\mathbf{S}$. The identifiable parameter combinations are summarized by the parameters of an alternative representation of the model known as the reduced-form VAR, which I describe next.

2.2 The Reduced-form Model

The structural model of the previous section implies the reduced-form model given by

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \mathbf{u}_t', \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \boldsymbol{\Sigma}), \quad \text{for } 1 \leq t \leq T, \tag{3}$$

[6] The non-identifiability of $\mathbf{S} \in \mathcal{S}^U$ is well known in the literature so I leave the formal proof to Appendix REF.

where the parameter matrices $(\mathbf{B}, \mathbf{\Sigma})$ and the forecast errors $\mathbf{u}_t$ are derived from the structural parameters as follows,

$$\mathbf{u}_t' = \boldsymbol{\varepsilon}_t'\mathbf{A}^{-1} \tag{4}$$

$$\mathbf{B} = \mathbf{F}\mathbf{A}^{-1} \tag{5}$$

$$\mathbf{\Sigma} = \left(\mathbf{A}\mathbf{A}'\right)^{-1} \tag{6}$$

and thus $\mathbf{B}$ and $\mathbf{\Sigma}$ have dimensions $(m \times n)$ and $(n \times n)$. Importantly, the matrix $\mathbf{\Sigma}$ is symmetric (and positive definite) and thus has only $n(n+1)/2$ unique elements. Similar to the notation for the structural model, define $\mathbf{D} \equiv (\mathbf{B}, \mathbf{\Sigma})$ where $\mathbf{D} \in \mathscr{D} = \mathbb{R}^{nm+n(n+1)/2}$, and $\mathcal{D} \equiv (p_{\mathbf{D}}(\cdot), \mathbf{D}, \mathscr{D})$, where $p_{\mathbf{D}}(\cdot)$ is the density for $\mathbf{y}_{1:T}$ implied by (3).

Although $\mathbf{S}$ are ultimately the objects of economic interest, the analysis of $\mathcal{D}$ often plays a key role in the practice of inferring $\mathbf{S}$. There are three reasons for this. First, despite the fact that $\mathscr{D}$ is of reduced dimensionality relative to $\mathscr{S}^U$, $\mathcal{D}$ and $\mathcal{S}$ are observationally equivalent models.[7] Second, $\mathbf{D}$ are globally identified while, as noted in the previous section, $\mathbf{S} \in \mathscr{S}^U$ are not.[8] The observational equivalence of the two models while only $\mathbf{D}$ are identified causes some researchers to describe the parameters $\mathbf{D}$ as summarizing the extent of the identifying information in the data. Under this interpretation one can see why many researchers would consider inference about $\mathbf{D}$ to be a natural starting place even if the aim is ultimately to infer $\mathbf{S}$. Third, the practical aspects of statistical inference are often more straightforward to implement for $\mathcal{D}$ than for $\mathcal{S}$ and hence, in practice, inference in $\mathcal{S}$ often proceeds by first estimating $\mathbf{D}$ and then mapping the estimate $\hat{\mathbf{D}}$ into an estimate $\hat{\mathbf{S}}$. In this discussion it would be difficult to overemphasize the importance of the observational equivalence of the two models, since it is that property that justifies the use of the estimation algorithms that first infer $\hat{\mathbf{D}}$.

The observational equivalence of the structural parameters under orthogonal rotations is closely related to the fact that only the reduced-form is identifiable.

---

[7] I consider two models $\mathcal{A}$ and $\mathcal{B}$ to be observationally equivalent if and only if for any $\theta_{\mathcal{A}} \in \Theta_{\mathcal{A}}$ there exists a $\theta_{\mathcal{B}} \in \Theta_{\mathcal{B}}$ for which $p(\mathbf{y}_{1:T}|\theta_{\mathcal{A}}) = p(\mathbf{y}_{1:T}|\theta_{\mathcal{B}})$ (and vice versa).

[8] The global identifiability of $\mathbf{D}$ is well-known in the literature. See, for example, Rothenberg (1971).

The right-hand multiplication by an orthogonal matrix $\mathbf{Q}$ of each element in a tuple will play a prominent role in the remainder of the paper, so I define the notation $\mathbf{S} * \mathbf{Q} \equiv (\mathbf{AQ}, \mathbf{FQ})$ and let $\mathcal{O}_n$ denote the set of $n \times n$ orthogonal matrices. It turns out that $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ are observationally equivalent only if there exists a $\mathbf{Q} \in \mathcal{O}_n$ such that $\widetilde{\mathbf{S}} = \mathbf{S} * \mathbf{Q}$. The "only if" part of this statement comes from the requirement to satisfy the relationship in equation (6) and a well known theorem guarantees that alternative full-rank square roots of an SPD matrix differ by only an orthogonal rotation; hence, the set of observationally equivalent structural parameters is the same as the set of structural parameters that yield identical values for $\mathbf{D}$.[9] I denote the set of structural parameters observationally equivalent to a given $\mathbf{D}$ as $\mathcal{S}(\mathbf{D})$. For any $\mathbf{y}_{1:T}$ and inferred $\mathbf{D}$, one might then say that there are as many possible SVARs hiding inside the reduced-form VAR as there are unique matrices in $\mathcal{O}_n$.

2.3 Identifying Restrictions: From Reduced-Form Estimation to Structural Inference

If a researcher begins with inference for $\mathbf{D}$ and wishes to progress to inference for $\mathbf{S}$ then he has two options for how to proceed. The first is to simply report the full set $\mathcal{S}(\hat{\mathbf{D}})$ for an estimated $\hat{\mathbf{D}}$. Choice two is to add more assumptions in the form of additional restrictions on the space of values for $\mathbf{S}$ considered fair game. Since every $\mathbf{S} \in \mathcal{S}(\hat{\mathbf{D}})$ is observationally equivalent, ruling out some elements of this set amounts to the imposition of what one might consider *classical* identifying restrictions, in the sense that the data cannot bear on their correctness.

When choosing the second option, let R denote some restrictions that a candidate $\mathbf{S}$ must satisfy and let $\mathcal{S}^R$ denote the space all structural parameters satisfying R. Rubio-Ramírez et al. (2010) show that, for a popular class of identifying restrictions, the statement that a VAR is exactly identified by restrictions R is equivalent to the statement that for almost every $\mathbf{S} \in \mathcal{S}^U$, there exists a *unique* matrix $\mathbf{Q} \in \mathcal{O}_n$ such that $\mathbf{S} * \mathbf{Q} \in \mathcal{S}^R$.[10] Presuming one has a way to find the

[9]For example, see Muirhead (1982) Theorem A9.5 which guarantees the existence of a $\mathbf{Q}$ in order for $\mathbf{AA}' = \tilde{\mathbf{A}}\tilde{\mathbf{A}}'$.

[10]See Theorem 5 of Rubio-Ramírez et al. (2010).

unique $\mathbf{Q}$, which will be a function of both $\mathbf{S}$ and R, inference for $\mathbf{S}$ in a restricted model can proceed via the following three-step procedure:

1. produce an estimate $\hat{\mathbf{D}}$,
2. map $\hat{\mathbf{D}}$ to an arbitrary $\tilde{\mathbf{S}} \in \mathcal{S}(\hat{\mathbf{D}})$, and[11]
3. construct the estimate of $\mathbf{S}$ as $\hat{\mathbf{S}} = \tilde{\mathbf{S}} * \mathbf{Q}(R, \tilde{\mathbf{S}})$.

Critical for the practicality of such an approach is that Rubio-Ramírez et al. (2010) provide an algorithm for finding the requisite $\mathbf{Q}$ to accompany an arbitrary $\mathbf{S} \in \mathcal{S}(\mathbf{D})$.[12] Critical for the legitimacy of this algorithm is the fact that $\hat{\mathbf{S}}$ and $\tilde{\mathbf{S}}$ have the same density under the prior of the unrestricted model, in addition to giving the same value to the data density.

In recent years researchers have found it desirable to investigate what can be learned when imposing even less than exactly identifying restrictions, a setting known as partial identification. Under partial identification $g_R^{-1}(\cdot)$ maps to a set and hence there is not a unique $\mathbf{Q}$ for which $\mathbf{S} * \mathbf{Q} \in \mathcal{S}^R$. When working within the Bayesian paradigm, the natural next step is to place a prior over the matrices in $\mathbf{Q}$. Rubio-Ramírez et al. (2010) give an efficient algorithm for generating random draws of $\mathbf{Q}$ distributed according to a uniform distribution over $\mathcal{O}_n$ known as Haar measure, samples from which can provide the basis of an accept-reject algorithm for Bayesian inference over the structural parameters when using sign restrictions.[13]

[11] An element of $\mathbf{S} \in \mathcal{S}(\mathbf{D})$ can be easily constructed as $(\tilde{\mathbf{A}}, \tilde{\mathbf{F}}) = g_{\text{chol}}^{-1}(\hat{\mathbf{D}}) = (\text{chol}(\hat{\mathbf{\Sigma}})^{-1}, \hat{\mathbf{B}}\,\text{chol}(\hat{\mathbf{\Sigma}})^{-1})$, where $\text{chol}(\mathbf{\Sigma})$ denotes the upper triangular Cholesky factor of an SPD matrix.

[12] See Algorithm 1 of Rubio-Ramírez et al. (2010).

[13] See Algorithm 2 in Rubio-Ramírez et al. (2010). Note that this procedure generates random draws consistent with only one the Haar measure distribution over $\mathcal{O}_n$. 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 partially identified SVARs.

2.4 Taking inventory of the SVAR's tractability

Before introducing my time-varying parameter framework, I pause to take inventory of the two key features of the constant parameter SVAR that make inference for $\mathbf{S}$ nearly trivial when R yields partial or exact identification: 1) the structural model is observationally equivalent to a tractable reduced-form and 2) one can easily map an $\hat{\mathbf{D}}$ to the (set of) $\mathbf{S}$ consistent with R In the next section I present my time-varying-parameter extension of S that mimics both features.

3. A Structural VAR with Time-Varying Parameters

In this section I describe my probabilistic framework for extending the structural model to allow for time-varying parameters. The purpose is to construct a class of "candidate truths" for the structural data generating process, among which the researcher aims to discriminate.

3.1 The Structural Model with Time-Varying Parameters

Allowing the structural matrices to change over time, the analogue to equation (2) in the time-varying parameter setting is

$$\mathbf{y}_t'\mathbf{A}_t = \mathbf{x}_t'\mathbf{F}_t + \boldsymbol{\varepsilon}_t', \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_{n\times 1}, \mathbf{I}_n), \quad \text{for } 1 \leq t \leq T\,. \tag{7}$$

However, it remains to specify the laws of motion for matrices of structural parameters. I specify the laws of motion such that the previous period's structural parameters are perturbed *multiplicatively* by a $(n \times n)$ matrix of random shocks $\mathbf{\Omega}_t$ and the previous period's lag coefficients $\mathbf{F}_{t-1}$ are also additively perturbed by an $(m \times n)$ matrix of mean-zero shocks $\mathbf{\Theta}_t$:

$$\mathbf{A}_t = \beta^{-1/2}\,\mathbf{A}_{t-1}\mathbf{\Omega}_t \tag{8}$$

$$\mathbf{F}_t = \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t + \mathbf{\Theta}_t \quad \text{for} \quad \mathbf{\Theta}_t \sim N(\mathbf{0}_{m\times n}, \mathbf{W}, \mathbf{I}_n) \tag{9}$$

where $\beta \in (n-1/n, 1)$ is a scalar and the notation $N(A, B, C)$ refers to a matrix-variate normal distribution with $m \times n$ mean matrix A, $m \times m$ row covariance matrix B, and $n \times n$ column covariance matrix C.

The parameter β has the interpretation of pulling the coefficients towards a loss of "information" in the system's implied reduced-form covariance matrix for shocks to $\mathbf{y}_t$. The old information is "replaced" by a random matrix $\mathbf{\Omega}_t$.

The structural shocks now consist of two distinct types of disturbances. The first type is the vector of structural shocks $\boldsymbol{\varepsilon}_t$, which perturb $\mathbf{y}_t$ through the equilibrium dynamics represented by $\mathbf{S}_t$, 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 affect objects of interests such as impulse response functions. The second type of shock perturbs the time t coefficients governing the equilibrium relationships among the variables, $\mathbf{S}_t$, and thus affect impulse responses. The random matrices $(\mathbf{\Omega}_t, \mathbf{\Theta}_t)$ are of this second type.

The random matrix $\mathbf{\Omega}_t$ is an orthogonal rotation from both left and right sides of a "square root" of a multivariate beta distributed random matrix, which is to say that it is realized according to

$$\mathbf{\Omega}_t = \mathbf{L}_t h(\mathbf{\Gamma}_t) \mathbf{R}_t \tag{10}$$

where $\mathbf{L}_t, \mathbf{R}_t \in \mathcal{O}_n$ and

$$\mathbf{\Gamma}_t \sim B_n(\nu(\beta)/2, 1/2) \tag{11}$$

for $B_n(\nu(\beta)/2, 1/2)$ denoting the n-dimensional multivariate beta distribution with degrees of freedom $\nu/2$ and $1/2$ as defined in Uhlig (1994).[14] The function $h(\cdot)$ returns the unique lower triangular Cholesky factor of an SPD matrix, with positive elements on the diagonal.[15] Thus $h(\cdot)$ maps $n \times n$ SPD matrices to a

[14]Traditionally the matrix beta distribution was defined only for $B_n(d_1/2, d_2/2)$ with $d_1, d_2 > n - 1$. Uhlig (1994) extends this definition to allow for d_2 to be a positive integer less than n, as in the definition of the innovations in (22). For further details also see Srivastava (2003). In some contexts the distribution is referred to as a "Type I" multivariate beta distribution.

[15]In principle, $h(\cdot)$ can be allowed to vary with t and each $h_t(\cdot)$ could be any function that returns a full rank factorization of its (full rank) matrix argument such that $h_t(\mathbf{\Gamma}_t)h_t(\mathbf{\Gamma}_t)' = \mathbf{\Gamma}_t$. However, since any two distinct factorizations of $\mathbf{\Gamma}_t$ can differ by only an orthogonal rotation, see Theorem A9.5. of Muirhead (1982), there is nothing lost by imposing that $h(\cdot)$ returns the

$n \times n$ matrices with $n(n+1)/2$ functionally independent elements. [16]

Before deriving the key properties of the DSVAR, I fix the following notation. I define $\boldsymbol{\phi} \equiv (\beta, \mathbf{W})$ to collect the model's static parameters and denote their space of possible values as $\boldsymbol{\Phi}$. I also extend the notation of the previous section so that $\mathbf{S}_t \equiv (\mathbf{A}_t, \mathbf{F}_t)$ denotes the tuple of structural matrices at time t and $\mathbf{S}_{0:T} \equiv (\mathbf{A}_{0:T}, \mathbf{F}_{0:T})$ denotes the sequence of tuples of structural parameters from times $t = 0$ to $t = T$. I then extend the definition of the "$*$" operator so that $\widetilde{\mathbf{S}}_{0:T} = \mathbf{S}_{0:T} * \mathbf{Q}_{0:T}$ yields $\widetilde{\mathbf{S}}_t = \mathbf{S}_t * \mathbf{Q}_t$ for each t. Lastly, the structural model involves the sequences of orthogonal matrices $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$, which are not separately identifiable from $\mathbf{S}_{0:T}$. For this reason I will treat the choice of $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ as part of the definition of a particular model rather than as free parameters to be estimated. For the class of models presented in this section I then write $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ to refer to the DSVAR with unrestricted parameter space and the particular choices of the orthogonal matrices $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$.

The Bayesian researcher aims to characterize the posterior distribution of the model's unobservables conditional on the data, $p(\boldsymbol{\phi}, \mathbf{S}_{0:T} | \mathbf{y}_{1:T})$, which one can factor as

$$p(\boldsymbol{\phi}, \mathbf{S}_{0:T} | \mathbf{y}_{1:T}) = c \cdot p(\boldsymbol{\phi}, \mathbf{S}_0) p(\mathbf{S}_{1:T} | \boldsymbol{\phi}, \mathbf{S}_0)\, p(\mathbf{y}_{1:T} | \boldsymbol{\phi}, \mathbf{S}_0, \mathbf{S}_{1:T}) \tag{12}$$

where c is an integrating constant that does not depend on the value of $(\boldsymbol{\phi}, \mathbf{S}_{0:T})$. A key property of the DSVAR is that a given $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ is observationally equivalent to a class of alternative models with different choices of $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$. I state the result formally, in the context of Bayesian inference, as follows

Theorem 1. *Let* $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ *have prior* $p(\boldsymbol{\phi}, \mathbf{S}_0)$ *for which* $p(\boldsymbol{\phi}, \mathbf{S}_0) =$

Cholesky factor and does not vary with time since the role of an alternative rotation can be absorbed into the choice of $\mathbf{R}_t$.

[16] Note that the framework can easily be extended to allow $\mathbf{F}_t$ to evolve according to

$$\mathbf{F}_t = \mathbf{G}\mathbf{F}_{t-1}\boldsymbol{\Omega}_t + \boldsymbol{\Theta}_t$$

for $\mathbf{G}$ a $(m \times m)$ matrix of static parameters. I focus on the case of of $\mathbf{G}$ dogmatically fixed at $\mathbf{I}_m$ because it most closely resembles the work in the VAR-TVP-SV literature and because researchers who have estimated related models with more flexible forms for $\mathbf{G}$ have typically found estimates that are close to the identity matrix. See, for example, Sims (1993).

$p(\boldsymbol{\phi}, \mathbf{S}_0 * \mathbf{P})$ for any $\mathbf{P} \in \mathcal{O}_n$. For any $\mathbf{Q}_{0:T}$ such that $\mathbf{Q}_t \in \mathcal{O}_n$, the model $\mathcal{S}^U_{0:T}(\widetilde{\mathbf{L}}_{1:T}, \widetilde{\mathbf{R}}_{1:T})$ defined by $(\widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t) = (\mathbf{Q}'_{t-1}\mathbf{L}_t, \mathbf{R}_t\mathbf{Q}_t)$ is such that, for every point $\mathbf{S}_{0:T}$, the point $\widetilde{\mathbf{S}}_{0:T} = \mathbf{S}_{0:T} * \mathbf{Q}_{0:T}$ satisfies

$$p\left(\boldsymbol{\phi}, \mathbf{S}_{0:T} | \mathbf{y}_{1:T}, \mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})\right) = p\left(\boldsymbol{\phi}, \widetilde{\mathbf{S}}_{0:T} | \mathbf{y}_{1:T}, \mathcal{S}^U_{0:T}(\widetilde{\mathbf{L}}_{1:T}, \widetilde{\mathbf{R}}_{1:T})\right) .$$

Proof. See Appendix A for the proof. □

Theorem 1 says that regardless of the realization of the observables one can always consider orthogonal rotations of the model's dynamic parameters and there will exist an alternative model justifying the choice of the rotated parameters. In the context of a full posterior density over values for $\mathbf{S}_{0:T}$, the relationship holds at every point under the same choice of $\mathbf{Q}_{0:T}$, and so the posterior of the model $\mathcal{S}^U_{0:T}(\widetilde{\mathbf{L}}_{1:T}, \widetilde{\mathbf{R}}_{1:T})$ is identical to that of $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ with each point in the parameter space rotated by $\mathbf{Q}_{0:T}$. Importantly, the specification of the alternative model does not depend on the realization of the data, but rather the relationship between the alternative model and parameter points is intrinsic to the model's definition. An obvious implication of Theorem 1 is that one cannot differentiate among different structural models on the basis of $\mathbf{y}_{1:T}$.

That such an invariance relation would hold for each of the densities $p(\mathbf{y}_t|\mathbf{S}_t, \boldsymbol{\phi})$ considered in isolation is not surprising since they have the same form as in the constant parameter model. Rather, the critical and surprising element of Theorem 1 lies in the fact that an invariance result in my model also holds for the density of the *sequence* of $\mathbf{S}_{0:T}$.

One can interpret Theorem 1 as defining a class of models observationally equivalent to a given $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$. Namely the models with

$$\widetilde{\mathbf{L}}_{1:T} = \mathbf{Q}'_{0:T-1} * \mathbf{L}_{1:T} \tag{13}$$

$$\widetilde{\mathbf{R}}_{1:T} = \mathbf{R}_{1:T} * \mathbf{Q}_{1:T} \tag{14}$$

for some $\mathbf{Q}_{0:T}$ with $\mathbf{Q}_t \in \mathcal{O}_n$ for each t. One might then say that there are as many models observationally equivalent to $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ as there are sequences of $T + 1$ orthogonal matrices.

As in the constant-parameter SVAR, the DSVAR implies a reduced-form model that is highly tractable for estimation, which can be profitably deployed in an algorithm for structural inference.

3.2 The Reduced-Form Model with Time-Varying Parameters

Multiplying both sides of equation (7) by $\mathbf{A}_t^{-1}$ gives the time-varying analogue to equation (3) as

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B}_t + \mathbf{u}_t', \quad \mathbf{u}_t \sim N(\mathbf{0}_{n\times 1}, \mathbf{\Sigma}_t), \quad \text{for } 1 \leq t \leq T. \tag{15}$$

where

$$\mathbf{u}_t' = \boldsymbol{\varepsilon}_t'\mathbf{A}_t^{-1}. \tag{16}$$

$$\mathbf{B}_t = \mathbf{F}_t\mathbf{A}_t^{-1} \tag{17}$$

$$\mathbf{\Sigma}_t = \left(\mathbf{A}_t\mathbf{A}_t'\right)^{-1} \tag{18}$$

and let $\mathbf{H}_t = \mathbf{\Sigma}_t^{-1}$ denote the precision matrix of the shocks $\mathbf{u}_t$. The definitions in equations (16)–(18) obviously resemble the relationship between the reduced-form VAR and SVAR in the constant-parameter case, but they are not sufficient to have fully specified a reduced-form TVP model. In order to make likelihood-based inference about the reduced-form implied by the DSVAR, laws of motion for the reduced-form parameters, consistent with the laws of motion for the underlying structural model, must be derived. I provide such laws of motion in this section.

I first derive the density for the reduced-form parameters, as defined in equations (16)–(18), under a given structural model $\mathcal{S}_{0:T}^U(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$. Denote $g(\mathbf{S}_t) = (\mathbf{A}_t\mathbf{A}_t', \mathbf{F}_t\mathbf{A}_t^{-1}) = (\mathbf{H}_t, \mathbf{B}_t)$ the density of the sequence $(\mathbf{H}_{0:T}, \mathbf{B}_{0:T})|\boldsymbol{\phi}$ can be factored as

$$p(\mathbf{H}_{0:T}, \mathbf{B}_{0:T}|\phi) = p(g(\mathbf{S}_0)|\boldsymbol{\phi}) \cdot \prod_{t=1}^{T} p(g(\mathbf{S}_t)|\boldsymbol{\phi}, \mathbf{S}_{0:t-1}) \tag{19}$$

$$= p(g(\mathbf{S}_0)|\boldsymbol{\phi}) \cdot \prod_{t=1}^{T} p(g(\mathbf{S}_t)|\boldsymbol{\phi}, \mathbf{S}_{t-1}) \,. \tag{20}$$

Each of the constituent densities in equation (20) can be further factored as

$$\begin{aligned} &p(\mathbf{A}_t\mathbf{A}_t', \mathbf{F}_t\mathbf{A}_t^{-1}|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}) \\ &\qquad = p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}) \cdot p(\mathbf{F}_t\mathbf{A}_t^{-1}|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t\mathbf{A}_t') \end{aligned} \tag{21}$$

Substituting into equations (17) and (18) with the definitions of the time t structural parameters based on their $t-1$ values and time t shocks gives the laws of motion for the reduced-form dynamic parameters as

$$\mathbf{H}_t = \mathbf{A}_t\mathbf{A}_t' = \frac{1}{\beta}\left(\mathbf{A}_{t-1}\mathbf{L}_t h(\boldsymbol{\Gamma}_t)\mathbf{R}_t\right)\left(\mathbf{R}_t' h(\boldsymbol{\Gamma}_t)'\mathbf{L}_t'\mathbf{A}_{t-1}'\right) \tag{22}$$

$$\mathbf{B}_t = \mathbf{F}_t\mathbf{A}_t^{-1} = \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1} + \boldsymbol{\Theta}_t\left(\beta^{-1/2}\mathbf{A}_{t-1}\mathbf{L}_t h(\boldsymbol{\Gamma}_t)\mathbf{R}_t\right)^{-1} \,, \tag{23}$$

where the definitions of, and distributions for, the random matrices $\boldsymbol{\Gamma}_t$ and $\boldsymbol{\Theta}_t$ are the same as in the previous section.[17] One can show that the densities in equation (21) take the form

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{A}_{t-1}) = p_{B_n}(\beta\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\mathbf{L}_t) \cdot |\beta^{1/2}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}|^{n+1} \,, \tag{24}$$

and

$$p(\mathbf{F}_t\mathbf{A}_t^{-1}|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t\mathbf{A}_t') = p_{MN}(\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}, \mathbf{W}, (\mathbf{A}_t\mathbf{A}_t')^{-1}) \,. \tag{25}$$

where the last term in equation (24) is the Jacobian.

The explicit dependence of the densities in equations (24) and (25) on the

[17]The expression in equation (23) is already partially simplified from $\mathbf{B}_t = \mathbf{F}_t\mathbf{A}_t^{-1} = \left(\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t + \boldsymbol{\Theta}_t\right)\mathbf{A}_t^{-1}$.

structural parameters in time $t-1$ would seem to suggest the problematic property that estimating the reduced-form parameters would require choosing a particular structural model $\mathcal{S}^{U}_{0:T}(\mathbf{L}_{1:T},\mathbf{R}_{1:T})$ and inferring $\mathbf{S}_{0:T}$ as we go. Such a property would be problematic in so much as it would mean that the reduced-form is no more tractable than the structural model itself and, if $\mathbf{S}_{0:T}$ are the objects of interest, one had might as well eschew the reduced-form model all together.[18]

Critically, I show that this is, in fact, not the case. First, note that the density of $(\mathbf{H}_{0:T},\mathbf{B}_{0:T})$ is invariant to the structural model's choice of $(\mathbf{L}_{1:T},\mathbf{R}_{1:T})$, which I formalize in the following result

Lemma 1. *The density of* $(\mathbf{H}_{0:T},\mathbf{B}_{0:T})$ *implied by a structural model* $\mathcal{S}^{U}_{0:T}(\mathbf{L}_{1:T},\mathbf{R}_{1:T})$ *is such that for any point* $(\mathbf{H}_{0:T},\mathbf{B}_{0:T})$

$$p(\mathbf{H}_{0:T},\mathbf{B}_{0:T}|\boldsymbol{\phi},\mathcal{S}^{U}_{0:T}(\mathbf{L}_{1:T},\mathbf{R}_{1:T})) = p(\mathbf{H}_{0:T},\mathbf{B}_{0:T}|\boldsymbol{\phi},\mathcal{S}^{U}_{0:T}(\widetilde{\mathbf{L}}_{1:T},\widetilde{\mathbf{R}}_{1:T}))\,.$$

for any $(\widetilde{\mathbf{L}}_{1:T},\widetilde{\mathbf{R}}_{1:T})$ *such that* $\widetilde{\mathbf{L}}_t,\widetilde{\mathbf{R}}_t\in\mathcal{O}_n$.

Proof. See Appendix.

Note that the laws of motion are such that the density of $\mathbf{B}_t$ conditions on the realization of $\mathbf{A}_t\mathbf{A}_t'$ and hence does not depend on $(\mathbf{L}_{1:T},\mathbf{R}_{1:T})$.

The key claim is then that $p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi},\mathbf{A}_{t-1})$ is invariant to $(\mathbf{L}_t,\mathbf{R}_t)$ □

The upshot of Lemma 1 is that alternative structural models imply the same densities for the sequence of reduced-form parameters. This result has two key implications. First, it suggests that, contrary to what one might have feared, the densities for $\mathbf{D}_{0:T}$ cannot truly depend on the particular $\mathbf{S}_{0:T}$ (else they would have been affected by the choice of structural model). Second, if one could estimate the reduced-form parameters directly, and we knew they were consistent with at least one structural model, then we would know that we were faithful to the reduced-form dynamics implied by an entire class of DSVAR models.

Building on Lemma 1, one can specify laws of motion that imply identical densities to those in equations (24) and (25), but which make no recourse to $\mathbf{S}_{0:T}$

[18] In such a setting one might also reasonably say that the "reduced-form" is a reparameterization rather than a true "reduced-form."

at all and which instead depend only on the sequence of reduced-form parameters. I formalize this result as follows:

Theorem 2. *Let $\mathcal{D}_{0:T}$ denote the model defined by the law of motion for observables $\mathbf{y}_{1:T}$ in equation* (15) *and by the following laws of motion for* $(\mathbf{H}_t, \mathbf{B}_t)$ *for all t*

$$\mathbf{H}_t = \frac{1}{\beta} h(\mathbf{H}_{t-1})\boldsymbol{\Gamma}_t h(\mathbf{H}_{t-1})' \quad \text{for} \quad \boldsymbol{\Gamma}_t \sim B_n(\nu(\beta)/2, 1/2) \tag{26}$$

$$\mathbf{B}_t = \mathbf{B}_{t-1} + \mathbf{V}_t \qquad\qquad \text{for} \quad \mathbf{V}_t \sim N(\mathbf{0}_{m\times n}, \mathbf{W}, \mathbf{H}_t^{-1}), \tag{27}$$

where $\nu(\beta) = \beta/(1-\beta)$.

If the prior for the structural model $p_{\mathcal{S}_0}(\mathbf{A}_0, \mathbf{F}_0)$ is such that $p_{g,\mathcal{S}_0}(g(\mathbf{A}_0, \mathbf{F}_0)) = p_{g,\mathcal{S}_0}(g(\mathbf{A}_0\mathbf{Q}_0, \mathbf{F}_0\mathbf{Q}_0))$ for any $\mathbf{Q}_0 \in \mathcal{O}_n$, and if the prior for $(\mathbf{H}_0, \mathbf{B}_0)$ is defined by $p_{\mathcal{D}_0}(\mathbf{H}_0, \mathbf{B}_0) = p_{g,\mathcal{S}_0}(g(\mathbf{A}_0, \mathbf{F}_0))$, then for any $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$,

$$p_{\mathcal{D}_{0:T}}(\mathbf{H}_{0:T}, \mathbf{B}_{0:T}|\boldsymbol{\phi}) = p_{g,\mathcal{S}_{0:T}}(\{g(\mathbf{A}_t, \mathbf{F}_t)\}_{t=1}^{T}|\boldsymbol{\phi}, \mathbf{L}_{1:T}, \mathbf{R}_{1:T})$$

Proof. See Appendix A.8. □

Theorem 2 has three key implications, the first of which I formalize in the following corollary.

Lemma 2. *For any $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$,*

$$p(g(\mathbf{S})|\boldsymbol{\phi}, \mathbf{S}_{t-1}, \mathbf{L}_t, \mathbf{R}_t) \propto p(\mathbf{S}_t|\boldsymbol{\phi}, \mathbf{S}_{t-1}, \mathbf{L}_t, \mathbf{R}_t) \tag{28}$$

Lemma 3.

$$p_{D_t}(\mathbf{D}_t|\boldsymbol{\phi}, \mathbf{D}_{t-1}) \propto p_{S_t}(\mathbf{S}_t|\boldsymbol{\phi}, \mathbf{S}_{t-1}) \tag{29}$$

Corollary 1. *$\mathcal{D}_{0:T}$ is a reduced-form of $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ for any $\mathbf{L}_{1:T}$ and $\mathbf{R}_{1:T}$.*

Proof. Each $\mathbf{B}_t$ and $\mathbf{F}_t$ have $m \times n$ unique unobservables. However, each $\mathbf{H}_t$ has only $n(n+1)/2$ unique unobservables while each $\mathbf{A}_t$ has n^2. Inspection of equations (15), (26), and (27) shows no residual dependence of $\mathbf{D}_t$ on the particular values of $\mathbf{S}_t$. The corollary then follows from Theorem 2 and the definition of a reduced-form of a structural model. $\square$

The second key implication of Theorem 2 is that, if one can estimate $\mathbf{D}_t$, then one can begin the task of structural inference by estimating $\mathcal{D}_{0:T}$ without taking a stand on a particular $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$.

Taken on its own, the model $\mathcal{D}_{0:T}$ is known in the statistics literature as a dynamic linear model with discounted-Wishart stochastic volatility (DLM-DWSV).

The $\mathcal{B}_n(d_1/2, d_2/2)$-distributed shocks have $\mathrm{E}[\boldsymbol{\Gamma}_t] = d_1/(d_1 + d_2)\,\mathbf{I}_n$. I will later assume that $d_1 = \beta/(1-\beta)$ and $d_2 = 1$, in which case $\mathrm{E}[\boldsymbol{\Gamma}_t] = \beta\,\mathbf{I}_n$. Given the linearity of expecations, one can see that the process for $\mathbf{H}_t$ then inherits random walk behavior.[19] The shocks to the VAR's linear coefficients $\mathbf{V}_t$ follow a matrix-normal distribution whose covariance matrix depends partially on the time t value of $\mathbf{H}_t^{-1}$.

An attractive feature of the DLM-DWSV is that it nests the constant-parameter reduced-form VAR as a limiting case. As $\beta \to 1$, the second moments of $\boldsymbol{\Gamma}_t$ collapse around the expectation and $\boldsymbol{\Gamma}_t \to \mathbf{I}_n$ for all t, so that $\mathbf{H}_t = \mathbf{H}_{t-1} = \cdots = \mathbf{H}_0$ thus eliminating the model's stochastic volatility component. Setting $\mathbf{G} = \mathbf{I}_m$ and taking $\mathbf{W} \to 0_{m\times m}$ implies that $\mathbf{B}_t = \mathbf{B}_{t-1} = \cdots = \mathbf{B}_0$, thus eliminating the model's TVP component. Hence, when we turn to the details of Bayesian estimation of the model the amount of shrinkage of the priors on β and $\mathbf{W}$ towards 1 and $0_{m\times m}$ respectively have the interpretation of controlling the size of the deviation from a constant-parameter VAR.

3.3 From DLM-DWSV Estimation back to Structural Inference

Given inference for $(\mathbf{D}_{0:T}, \boldsymbol{\phi})$, inference for $(\mathbf{S}_{0:T}, \boldsymbol{\phi})$ can then proceed in much the same way as in constant-parameter models. Choices of $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ map $\mathbf{D}_{0:T}$ into $\mathbf{S}_{0:T}$ and thus amount to choosing among the observationally equivalent DSVARs hiding inside of the estimated DLM-DWSV.

[19]Note that $\mathrm{E}[\mathbf{H}_t|\mathbf{H}_{t-1}] = \frac{1}{\beta}h(\mathbf{H}_{t-1})(\beta\mathbf{I}_n)h(\mathbf{H}_{t-1}) = \mathbf{H}_{t-1}$.

As a first step, an arbitrary $\widetilde{\mathbf{S}}_{0:T}$ can be constructed from an estimate $\widehat{\mathbf{D}}_{0:T}$ as

$$\widetilde{\mathbf{A}}_t = h(\widehat{\mathbf{H}}_t) \tag{30}$$

$$\widetilde{\mathbf{F}}_t = \widehat{\mathbf{B}}_t\, h(\widehat{\mathbf{H}}_t)\,. \tag{31}$$

If one then aims to restrict attention to $\mathbf{S}_{0:T}$ for which each $\mathbf{S}_t \in \mathscr{R}_t$, then one can use the elements of $\widetilde{\mathbf{S}}_{0:T}$ as inputs into the period-by-period application of either Algorithm 1 of Rubio-Ramírez et al. (2010) for the case where $\mathscr{R}_t$ yields exact identification, or Algorithm 2 of Rubio-Ramírez et al. (2010) for the case where $\mathscr{R}_t$ yields partial identification.

4. Bayesian Estimation

In this section I describe my highly tractable MCMC algorithm for the fully-Bayesian estimation of all reduced-form model parameters. After having estimated $\boldsymbol{\phi}, \mathbf{D}_{0:T} | \mathbf{y}_{1:T}, \mathcal{D}$, inference for structural parameters can then proceed as outlined in Section 3.3.

4.1 Bayesian Inference for Reduced-form Parameters

Given a sample of data $\mathbf{y}_{1:T}$, the goal is to characterize the posterior distribution of the model's unobservables:

$$p(\boldsymbol{\phi}, \mathbf{D}_{0:T} | \mathbf{y}_{1:T}) = \frac{p(\boldsymbol{\phi}, \mathbf{D}_{0:T}) p(\mathbf{y}_{1:T} | \boldsymbol{\phi}, \mathbf{D}_{0:T})}{p(\mathbf{y}_{1:T})} \tag{32}$$

One cannot fully characterize the posterior analytically so I propose to make inference about $(\boldsymbol{\phi}, \mathbf{D}_{0:T})$ 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, sampling iteratively will yield draws representative of the model's posterior asymptotically in the length of the chain. In particular, my MCMC algorithm is of type known as a Gibbs Sampler, which means that the algorithm entails iteratively sampling from the *conditional* posteriors of different "blocks" of a partition of the model's unobservables.

I now sketch the key elements of the Gibbs sampler, leaving the details and exact formulas for Appendix C.[20] My Gibbs sampler for the DLM-DWSV consists of two main blocks of parameters based on the partition of the unobservables into $\mathbf{W}$ and $(\beta, \mathbf{D}_{0:T})$, which involves sampling from the following sequence of conditional distributions,

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

The draw from Block 1 is straigtforward, while the draw from Block 2 is more nuanced. Under an Inverse Wishart prior for $\mathbf{W}$, the conditional posterior in Block 1 is also an Inverse Wishart distribution. In Block 2 I sample from the joint posterior of $\beta, \mathbf{D}_{0:T}|\mathbf{y}_{1:T}, \mathbf{W}$ by means of factoring the joint distribution into the distribution of $\beta|\mathbf{y}_{1:T}, \mathbf{W}$ in Step 2(a), which is *marginal* of $\mathbf{D}_{0:T}$, and the distribution of $\mathbf{D}_{0:T}|\mathbf{y}_{1:T}, \mathbf{W}, \beta$ in Step 2(b), which conditions on the value of β.

The feasability of my sampling strategy for Block 2 hinges on 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 $\mathbf{D}_{0:T}$, *including the stochastic volatility components*, in a fashion analogous to likelihood-based inference with the Kalman filter. The draw of $\beta|\mathbf{y}_{1:T}, \mathbf{W}$ in Step 2(a) can then be implemented with a so-called "Metropolis-within-Gibbs" step since one can evaluate the likelihood. I use a standard random-walk Metropolis-Hastings algorithm. Second, there exist exact expressions for recursively sampling backwards a sequence of latent states from their conditional posterior (sometimes known as a smoother), which is what occurs in Step 2(b).

[20] Appendix D contains further considerations pertinent to the computationally efficient implementation of the algorithms.

4.2 Priors

This section describes the general structure of some useful classes of priors with the specific choices of prior hyperparameters provided in the context of the application. The model primitives requiring prior distributions are β,$\mathbf{W}$, and $\mathbf{D}_0$.

Prior for β. It's obvious from the "discounting" interpretation of β's role in the model that one must at least restrict its support to [0, 1], however there are other considerations one must keep in mind as well. Furthermore, from Table A-1 one can see that there will be Wishart distributions characterizing our uncertainty over the $n \times n$ matrix H_t with degrees of freedom given by βh_{t-1} and $\beta h_{t-1} + 1$. 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 valid probability distributions at each step we then need β to be such that $\beta/(1-\beta) > n-1$, which implies that we need restrict the postive density of $p(\beta)$ to $\beta > (n-1)/n$. In applications 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.

Prior for W. In the VAR-TVP-SV model of Primiceri (2005) the distribution of the model's linear coefficients is

$$\mathbf{b}_t \sim \mathcal{N}(\mathbf{b}_{t-1}, Q) \; . \tag{33}$$

Beginning with Primiceri (2005), it has become standard in the VAR-TVP-SV literature to base the prior for Q on a pre-sample of observations. The prior in Primiceri (2005) takes the form

$$Q \sim \mathcal{IW}(k_Q^2 \cdot T_{pre} \cdot V(\hat{B}_{OLS}),\ T_{pre}) \tag{34}$$

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.[21] In Primiceri (2005), $k_Q = 0.01$ and $T_{pre} = 40$ and $V(\hat{B}_{OLS}) = \Sigma_{pre} \otimes (X'_{pre} X_{pre})^{-1}$.

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

In the DLM-DWSV, $\mathbf{b}_t$ has distribution

$$\mathbf{b}_t \sim \mathcal{N}(\mathbf{b}_{t-1}, \boldsymbol{\Sigma}_t \otimes \mathbf{W}) \tag{35}$$

and hence the matrix $(\boldsymbol{\Sigma}_t \otimes \mathbf{W})$ functions similarly to Q from the TVP-VAR-SV. I choose a prior for $\mathbf{W}$ in the spirit of the $p(Q)$ given in (34). Estimating a VAR over a presample under a diffuse prior yields the posterior for $\mathbf{b}_{pre}$ of

$$\mathbf{b}_{pre} | \Sigma_{pre} \sim \mathcal{N}(\hat{b}_{pre}, \Sigma_{pre} \otimes (X'_{pre} X_{pre})^{-1}) \, . \tag{36}$$

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

$$W \sim \mathcal{IW}(\delta_1^2 \cdot T_{pre} \cdot (X'_{pre} X_{pre})^{-1}, T_{pre}) \, . \tag{37}$$

Prior for dynamic latent states. The prior for the initial values of the dynamic latent states $(\mathbf{H}_0, \mathbf{B}_0)$ maintains the form of the distributional families used in the recursive filter summarized in Table A-1, that is $(\mathbf{H}_0,\ \mathbf{B}_0) \sim \mathcal{NW}(\mathbf{M}_0, \mathbf{C}_0, h_0, \mathbf{D}_0)$. In the context of the recursions described in Table A-1, one can think of this distribution as a posterior from $t = 0$, for which there was no observation from which to update the latent states. The prior for $(\mathbf{H}_1, \mathbf{B}_1)$ is then induced by the move from Step 0 to Step 1 as described in Table A-1. Treating $\mathbf{D}_0$ in this fashion allows its elements to be integrated out of the likelihood just like the rest of the sequence $\mathbf{D}_{1:T}$.

The remaining primitives to be specified by the researcher are then (M_0, C_0, h_0, D_0). My benchmark choices for these values are informed by the DLM-DWSV's property that certain limiting cases of $\boldsymbol{\phi}$ collapse the model to a standard VAR with a conjugate prior $\mathcal{NW}(M_0, C_0, h_0, D_0)$ for time invariant $\mathbf{D}$. As $\beta \to 1$, $\mathbf{H}_t \to \mathbf{H}_{t-1}$ for all t and hence the full sequence of H_t values collapses to H_0. From the recursions in Table A-1 one can see that $\beta \to 1$ also causes the degrees of freedom in the Wishart distribution accumulate from h_0 in one-for-one with the acquisition of new observations, as in standard VAR estimation. These two observations suggests a prior for (h_0, D_0) that moves closer to that of a standard

VAR as $\beta \to 1$. While smaller values of β imply greater variability over the sequence of $\mathbf{\Sigma}_t$ in which case one might want the prior for $\mathbf{H}_0$ to be more diffuse.

5. Application: Time-Varying Fiscal Multipliers Under Three Identification Schemes

I apply my DSVAR to the question of whether or not the fiscal multiplier varies over time, possibly with the state of the economy. The question of time-variation in the fiscal multiplier has engendered a significant resurgence in the wake of the Great Recession as the severity of the crisis left policy makers looking for beneficial interventions, particularly in light of the monetary authority having exhausted its standard tools. It has been argued on the basis of various theoretical models that the fiscal multiplier increases with the amount of "slack" in the economy or the nominal interest rate is constrained by the zero lower bound. Ramey and Zubairy (2017) and Auerbach and Gorodnichenko (2012) represent prominent attempts to address the question empirically from time-series data.

6. Conclusion

TO BE COMPLETED.

A. Proofs of Theorems

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 the TVP case.

Lemma 4. *If* $\mathbf{S}$ *and* $\tilde{\mathbf{S}}$ *are values of the structural parameters for* $\mathcal{S}^U$ *such that* $\widetilde{\mathbf{S}} = \mathbf{S} * \mathbf{Q}$ *for* $\mathbf{Q} \in \mathcal{O}_n$*, then* $p(\mathbf{y}_t | \mathbf{S}, \mathbf{y}_{t-p:t-1}) = p(\mathbf{y}_t | \widetilde{\mathbf{S}}, \mathbf{y}_{t-p:t-1})$.

Proof. First rewrite equation (2) as

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{F}\mathbf{A}^{-1} + \varepsilon_t'\mathbf{A}^{-1}\,, \tag{A.38}$$

where $\mathbf{y}_t$ thus has density

$$\begin{aligned} p(\mathbf{y}_t | \mathbf{S}, \mathbf{y}_{t-p:t-1}) &= (2\pi)^{-1/2} |(\mathbf{A}\mathbf{A}')^{-1}|^{-1/2} \\ &\quad \cdot \exp\left\{-(1/2) \cdot (\mathbf{y}_t' - \mathbf{x}_t'\mathbf{F}\mathbf{A}^{-1})(\mathbf{A}\mathbf{A}')(\mathbf{y}_t' - \mathbf{x}_t'\mathbf{F}\mathbf{A}^{-1})'\right\}\,. \end{aligned} \tag{A.39}$$

The lemma follows from evaluating equation (A.39) at the parameter point $\widetilde{\mathbf{S}}$ and noting that

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}' = \mathbf{A}\mathbf{Q}\mathbf{Q}'\mathbf{A}' = \mathbf{A}\mathbf{A}' \tag{A.40}$$

$$\tilde{\mathbf{F}}\tilde{\mathbf{A}}^{-1} = \mathbf{F}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{A}^{-1} = \mathbf{F}\mathbf{A}' \tag{A.41}$$

where the second equalities in equations (A.40) and (A.41) follow from the orthogonality of $\mathbf{Q}$. □

Corollary 2. *In the environment of Lemma 4,* $p(\mathbf{y}_{1:T} | \mathbf{S}) = p(\mathbf{y}_{1:T} | \widetilde{\mathbf{S}})$.

Proof.

$$p(\mathbf{y}_{1:T}|\mathbf{S}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{S}, \mathbf{y}_{0:t-1}) \tag{A.42}$$

$$= \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{S}, \mathbf{y}_{t-p:t-1}) \tag{A.43}$$

$$= \prod_{t=1}^{T} p(\mathbf{y}_t|\widetilde{\mathbf{S}}, \mathbf{y}_{t-p:t-1}) \tag{A.44}$$

$$= p(\mathbf{y}_{1:T}|\widetilde{\mathbf{S}})\,. \tag{A.45}$$

where the equality in (A.44) follows from Lemma 4. □

A.2 Densities of time-varying structural parameters

This section derives the density of $\mathbf{S}_t$ under the laws of motion for $\mathbf{A}_t$ and $\mathbf{F}_t$ in equations (8) and (9). One can factor each conditional density as

$$p(\mathbf{S}_t|\boldsymbol{\phi}, \mathbf{S}_{t-1}) = p(\mathbf{A}_t, \mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}) \tag{A.46}$$

$$= p(\mathbf{A}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}) \cdot p(\mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t) \tag{A.47}$$

$$= p(\mathbf{A}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}) \cdot p(\mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t) \tag{A.48}$$

where (A.56) follows from the definition of the law of motion for $\mathbf{A}_t$ in equation (8), which does not depend on $\mathbf{F}_{t-1}$.

Claim 1. *Under the model* $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$*,*

$$\begin{aligned} &p(\mathbf{A}_t|\boldsymbol{\phi},\mathbf{A}_{t-1}, \mathbf{L}_t, \mathbf{R}_t) \\ &\quad = p_{B_n}\left(\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\mathbf{L}_t\sqrt{\beta}\,\middle|\,\nu(\beta)/2, 1/2)\right) \\ &\qquad \cdot \prod_{i=1}^{n}\left[\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{R}_t'\right]_{ii}^{n-i} \cdot \beta^{n/2}\det(\mathbf{A}_{t-1}^{-1})^n\,, \end{aligned} \tag{A.49}$$

where $p_{B_n}(\cdot|\nu(\beta)/2, 1/2)$ *denotes the pdf of the singular multivariate beta distribution with degrees of freedom* $\nu(\beta)/2$ *and* $1/2$ *for* $\nu(\beta) = \beta/(1-\beta)$*.*

Proof. I provide the proof in Appendix F. □

Claim 2. *Under the model $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$,*

$$p(\mathbf{F}_t|\boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t, \mathbf{L}_t, \mathbf{R}_t) = p_N(\mathbf{F}_t|\overline{\mathbf{F}}_t, \mathbf{W}, \mathbf{I}_n) \tag{A.50}$$

for $\overline{\mathbf{F}}_t = \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t$, where $p_N(\cdot|A, B, C)$ denotes the pdf of the matrix-variate normal distribution with mean matrix A, row-covariance matrix B, and column-covariance matrix C.

Proof. Under the law of motion in equation (9), $\mathbf{F}_t = \overline{\mathbf{F}}_t + \boldsymbol{\Theta}_t$, the result is immediate from the definition of $\boldsymbol{\Theta}_t$ and well-known properties of the matrix-variate normal distribution. □

A.3 Proof of Theorem 1

The key element of the proof Theorem 1 is the following result regarding the joint density of the data and the dynamic unobservables, which I state and prove before proving Theorem 1.

Lemma 5. *If $\mathbf{S}_{0:T}$, $\widetilde{\mathbf{S}}_{0:T}$, $(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$, and $(\widetilde{\mathbf{L}}_{1:T}, \widetilde{\mathbf{R}}_{1:T})$ are as defined in Theorem 1, then*

$$\begin{aligned} &p\left(\mathbf{y}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0, \mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})\right) \\ &\qquad\qquad = p\left(\mathbf{y}_{1:T}, \widetilde{\mathbf{S}}_{1:T}|\boldsymbol{\phi}, \widetilde{\mathbf{S}}_0, \mathcal{S}^U_{0:T}(\widetilde{\mathbf{L}}_{1:T}, \widetilde{\mathbf{R}}_{1:T})\right) . \end{aligned}$$

Proof of Lemma 5. I prove the lemma in three parts.

Part I: Preliminaries.

First, factor the joint density $p(\mathbf{y}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0)$ as the product of a marginal and a conditional

$$p(\mathbf{y}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0) = p(\mathbf{y}_{1:T}|\mathbf{S}_{1:T}, \boldsymbol{\phi}, \mathbf{S}_0) \cdot p(\mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0) \tag{A.51}$$

$$= p(\mathbf{y}_{1:T}|\mathbf{S}_{1:T}) \cdot p(\mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0)\,, \tag{A.52}$$

where the second equality simply follows from the definition of the data density in (7). By Corollary 2, $p(\mathbf{y}_{1:T}|\mathbf{S}_{1:T}) = p(\mathbf{y}_{1:T}|\widetilde{\mathbf{S}}_{1:T})$. Hence, it remains only to show that $p(\mathbf{S}_{1:T}|\mathbf{S}_0, \boldsymbol{\phi}) = p(\widetilde{\mathbf{S}}_{1:T}|\widetilde{\mathbf{S}}_0, \boldsymbol{\phi})$.

By the Markovian assumption implicit in the definition of the laws of motion for $\mathbf{A}_t$ and $\mathbf{F}_t$ in equations (8) and (9), one can write the density for the sequence $\mathbf{S}_{1:T}$ as

$$p(\mathbf{S}_{1:T}|\boldsymbol{\phi}) = \prod_{t=1}^{T} p(\mathbf{S}_t|\mathbf{S}_{t-1}, \boldsymbol{\phi}) \tag{A.53}$$

and factor each conditional density in equation (A.53) as

$$p(\mathbf{S}_t|\mathbf{S}_{t-1}, \boldsymbol{\phi}) = p(\mathbf{A}_t, \mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}) \tag{A.54}$$
$$= p(\mathbf{A}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}) \cdot p(\mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}, \mathbf{A}_t) \tag{A.55}$$
$$= p(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}) \cdot p(\mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}, \mathbf{A}_t) \tag{A.56}$$

where (A.56) follows from the definition of the law of motion for $\mathbf{A}_t$ in equation (8), which does not depend on $\mathbf{F}_{t-1}$. One can thus rewrite equation (A.53) as

$$p(\mathbf{S}_{1:T}|\mathbf{S}_0, \boldsymbol{\phi}) = \left[\prod_{t=1}^{T} p(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi})\right] \cdot \left[\prod_{t=1}^{T} p(\mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}, \mathbf{A}_t)\right]. \tag{A.57}$$

To prove the theorem it will then suffice to show that the following two equalities hold

$$p(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t) = p(\tilde{\mathbf{A}}_t|\tilde{\mathbf{A}}_{t-1}, \boldsymbol{\phi}, \tilde{\mathbf{L}}_t, \tilde{\mathbf{R}}_t) \tag{A.58}$$
$$p(\mathbf{F}_t|\mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \boldsymbol{\phi}, \mathbf{A}_t) = p(\tilde{\mathbf{F}}_t|\tilde{\mathbf{A}}_{t-1}, \tilde{\mathbf{F}}_{t-1}, \boldsymbol{\phi}, \tilde{\mathbf{A}}_t), \tag{A.59}$$

for

$$(\tilde{\mathbf{A}}_t, \tilde{\mathbf{F}}_t, \tilde{\mathbf{A}}_{t-1}, \tilde{\mathbf{F}}_{t-1}) = (\mathbf{A}_t\mathbf{Q}_t, \mathbf{F}_t\mathbf{Q}_t, \mathbf{A}_{t-1}\mathbf{Q}_{t-1}, \mathbf{F}_{t-1}\mathbf{Q}_{t-1}), \tag{A.60}$$

and

$$(\tilde{\mathbf{L}}_t, \tilde{\mathbf{R}}_t) = (\mathbf{Q}_{t-1}^{-1}\mathbf{L}_t, \mathbf{R}_t\mathbf{Q}_t) \tag{A.61}$$

for any $\mathbf{Q}_t, \mathbf{Q}_{t-1}, \mathbf{L}_t, \mathbf{R}_t \in \mathcal{O}_n$. Hence, I proceed by showing that the equalities in (A.58) and (A.59) do indeed hold. Note that the argument that the relevant equalities hold is somewhat more nuanced than it might first appear because there are two "moving parts" in each expression on the right-hand side of equations (A.58) and (A.59). For example, conditioning on the value $\tilde{\mathbf{A}}_{t-1}$ in (A.58) instead of $\mathbf{A}_{t-1}$ changes the density of $\mathbf{A}_t$, and hence the new density must be derived and then evaluated at the parameter point $\tilde{\mathbf{A}}_t$.

Part II: Proof that equation (A.58) *holds.*

Consider a new random variable $\widehat{\mathbf{A}}_t$ with law of motion given by

$$\widehat{\mathbf{A}}_t = \beta^{-1/2}\widetilde{\mathbf{A}}_{t-1}\widetilde{\mathbf{\Omega}}_t \tag{A.62}$$

$$= \beta^{-1/2}\widetilde{\mathbf{A}}_{t-1}\widetilde{\mathbf{L}}_t h(\mathbf{\Gamma}_t)\widetilde{\mathbf{R}}_t \tag{A.63}$$

for $(\widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t)$ as defined in equation (A.61).

The following relationships will be used repeatedly,

$$\widetilde{\mathbf{L}}_t'\widetilde{\mathbf{A}}_{t-1}^{-1} = (\mathbf{L}_t'\mathbf{Q}_{t-1}')(\mathbf{Q}_{t-1}\mathbf{A}_{t-1}^{-1}) = \mathbf{L}_t'\mathbf{A}_{t-1}^{-1} \tag{A.64}$$

$$\widehat{\mathbf{A}}_t\widetilde{\mathbf{R}}_t' = \widehat{\mathbf{A}}_t\mathbf{Q}_t'\mathbf{R}_t' \tag{A.65}$$

when evaluated at $\widehat{\mathbf{A}}_t = \mathbf{A}_t\mathbf{Q}_t$, equation (A.65) becomes

$$\mathbf{A}_t\mathbf{Q}_t\mathbf{Q}_t'\mathbf{R}_t' = \mathbf{A}_t\mathbf{R}_t' \,. \tag{A.66}$$

I now substitute into the density for $\mathbf{A}_t$ term-by-term. Beginning with $p_\Gamma(\cdot)$,

$$p_{\mathbf{\Gamma}}\left(h^{-1}(g^{-1}(\widehat{\mathbf{A}}_t|\widetilde{\mathbf{A}}_{t-1}, \boldsymbol{\phi}))\right) = p_{B_n}(\sqrt{\beta}\widetilde{\mathbf{L}}_t'\widetilde{\mathbf{A}}_{t-1}^{-1}\widehat{\mathbf{A}}_t\widetilde{\mathbf{R}}_t'\widetilde{\mathbf{R}}_t\widehat{\mathbf{A}}_t'\widetilde{\mathbf{A}}_{t-1}^{-1\prime}\widetilde{\mathbf{L}}_t\sqrt{\beta}) \,. \tag{A.67}$$

Using the relationships in equations (A.64) and (A.65) and evaluating at the point

$\widehat{\mathbf{A}}_t = \mathbf{A}_t\mathbf{Q}_t$ gives

$$p_{\mathbf{\Gamma}}\left(h^{-1}(g^{-1}(\widehat{\mathbf{A}}_t|\widetilde{\mathbf{A}}_{t-1},\boldsymbol{\phi}))\right) = p_{B_n}(\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{R}_t'\mathbf{R}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\mathbf{L}_t\sqrt{\beta}) \,. \tag{A.68}$$

which matches the expression in equation (F.182).

Next turning to the Jacobian terms,

$$|J(h^{-1})| = \prod_{i=1}^{n}[h(\mathbf{\Gamma}_t)]_{ii}^{n-i} \tag{A.69}$$

$$= \prod_{i=1}^{n}\left[\sqrt{\beta}\widetilde{\mathbf{L}}_t'\widetilde{\mathbf{A}}_{t-1}^{-1}\widehat{\mathbf{A}}_t\widetilde{\mathbf{R}}_t'\right]_{ii}^{n-i} \tag{A.70}$$

$$= \prod_{i=1}^{n}\left[\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{R}_t'\right]_{ii}^{n-i} \,, \tag{A.71}$$

which matches the expression in equation (F.183) and where the last equality comes from simple substitutions and the expressions in (A.64).

Next considering $|Jg^{-1}|$,

$$|Jg^{-1}| = \underbrace{\det(\widetilde{\mathbf{R}}_t')^n}_{=1} \cdot \det(\sqrt{\beta}\widetilde{\mathbf{L}}_t'\widetilde{\mathbf{A}}_{t-1}^{-1})^n \tag{A.72}$$

$$= \beta^{n/2}\det(\mathbf{L}_t'\mathbf{A}_{t-1}^{-1})^n \tag{A.73}$$

which matches the expression in equation (F.186). Thus the proof of Part II is complete.

Part III: Proof that equation (A.59) *holds.*

The density is given by

$$\begin{aligned} p(\mathbf{F}_t|\mathbf{A}_{t-1},\mathbf{F}_{t-1},\boldsymbol{\phi},\mathbf{A}_t) &= (2\pi)^{-nm/2}|\mathbf{I}_n|^{m/2}|\mathbf{W}|^{n/2} \\ &\quad\cdot \exp\{-\frac{1}{2}\mathrm{tr}[\mathbf{I}_n^{-1}(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})'\mathbf{W}^{-1}(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})]\} \end{aligned} \tag{A.74}$$

Now consider the new random variable $\hat{\mathbf{F}}_t$ where

$$\hat{\mathbf{F}}_t = \tilde{\mathbf{F}}_{t-1}\tilde{\mathbf{A}}_{t-1}^{-1}\tilde{\mathbf{A}}_t + \mathbf{\Theta}_t \tag{A.75}$$

Substituting into equation (A.75) with the definitions of $\tilde{\mathbf{A}}_{t-1}, \tilde{\mathbf{F}}_{t-1}, \tilde{\mathbf{A}}_t$ gives

$$\hat{\mathbf{F}}_t = \mathbf{F}_{t-1}\left(\mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^{-1}\right)\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{Q}_t + \mathbf{\Theta}_t \tag{A.76}$$

$$= \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{Q}_t + \mathbf{\Theta}_t \tag{A.77}$$

$$= \mathbf{M}_{\mathbf{F},t}\mathbf{Q}_t + \mathbf{\Theta}_t \tag{A.78}$$

and hence

$$\hat{\mathbf{F}}_t|\tilde{\mathbf{A}}_{t-1}, \tilde{\mathbf{F}}_{t-1}, \boldsymbol{\phi}, \tilde{\mathbf{A}}_t \sim \mathcal{MN}(\mathbf{M}_{\mathbf{F},t}\mathbf{Q}_t,\ \mathbf{W},\ \mathbf{I}_n) \tag{A.79}$$

The density of $\hat{\mathbf{F}}_t$ and the density of $\mathbf{F}_t$ differ by only the arguments of the exponential-trace term of the matrix-variate normal density. The trace term from the density of $\hat{\mathbf{F}}_t$ under the distribution in (A.79) is

$$\operatorname{tr}\left[\mathbf{I}_n^{-1}(\hat{\mathbf{F}}_t - \mathbf{M}_{\mathbf{F},t}\mathbf{Q}_t)'\mathbf{W}^{-1}(\hat{\mathbf{F}}_t - \mathbf{M}_{\mathbf{F},t}\mathbf{Q}_t)\right]. \tag{A.80}$$

Evaluating (A.80) at the point $\hat{\mathbf{F}}_t = \tilde{\mathbf{F}}_t = \mathbf{F}_t\mathbf{Q}_t$ gives

$$\operatorname{tr}\left[\mathbf{I}_n^{-1}(\mathbf{F}_t\mathbf{Q}_t - \mathbf{M}_{\mathbf{F},t}\mathbf{Q}_t)'\mathbf{W}^{-1}(\mathbf{F}_t\mathbf{Q}_t - \mathbf{M}_{\mathbf{F},t}\mathbf{Q}_t)\right] \tag{A.81}$$

$$= \operatorname{tr}\left[\mathbf{I}_n^{-1}\mathbf{Q}_t'(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})'\mathbf{W}^{-1}(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})\mathbf{Q}_t\right] \tag{A.82}$$

$$= \operatorname{tr}\left[\mathbf{Q}_t\mathbf{I}_n^{-1}\mathbf{Q}_t'(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})'\mathbf{W}^{-1}(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})\right] \tag{A.83}$$

$$= \operatorname{tr}\left[\mathbf{I}_n^{-1}(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})'\mathbf{W}^{-1}(\mathbf{F}_t - \mathbf{M}_{\mathbf{F},t})\right] \tag{A.84}$$

where (A.83) and (A.84) follow from the cyclical property of the trace operator and the orthogonality of $\mathbf{Q}_t$. The expression in (A.84) matches the trace term in (A.74), which completes the proof. □

With Lemma 5 in hand, the desired result follows almost immediately.

Proof of Theorem 1. Using the factorization of the posterior in equation (12),

$$p(\boldsymbol{\phi}, \mathbf{S}_{0:T}|\mathbf{y}_{1:T}) = c \cdot p(\boldsymbol{\phi}, \mathbf{S}_0) \underbrace{p(\mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0)\, p(\mathbf{y}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0, \mathbf{S}_{1:T})}_{p(\mathbf{y}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0)},$$

the result follows from Lemma 5 ensuring the equality of $p(\mathbf{y}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\phi}, \mathbf{S}_0) = p(\mathbf{y}_{1:T}, \widetilde{\mathbf{S}}_{1:T}|\boldsymbol{\phi}, \widetilde{\mathbf{S}}_0)$, and the premise that $p(\boldsymbol{\phi}, \mathbf{S}_0) = p(\boldsymbol{\phi}, \widetilde{\mathbf{S}}_0)$. □

A.4 Densities of $g(\mathbf{S}_t)$

In this section I derive the density of the random variables defined by $g(\mathbf{A}_t, \mathbf{F}_t)$.

Claim 3.

$$p_{g,S_t}(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1}) = p_{B_n}(\beta \mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\mathbf{L}_t) \cdot \beta^{n(n+1)/2}|\det(\mathbf{A}_{t-1})|^{-(n+1)}.$$

Proof. See Appendix F. □

Claim 4. *For any* $\mathbf{L}_t, \mathbf{R}_t \in \mathcal{O}_n$,

$$p_{g,S_t}(\mathbf{F}_t\mathbf{A}_t^{-1}|\mathbf{F}_{t-1}, \mathbf{A}_{t-1}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_t) = p_N(\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}, \mathbf{W}, (\mathbf{A}_t\mathbf{A}_t')^{-1}).$$

Proof. Repeating the law of motion from the main text when conditioning on $\mathbf{A}_t$,

$$\text{(A.85)} \qquad \mathbf{F}_t\mathbf{A}_t^{-1} = \left(\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}\mathbf{A}_t + \boldsymbol{\Theta}_t\right)\mathbf{A}_t^{-1} \quad \text{for} \quad \boldsymbol{\Theta}_t \sim N(\mathbf{0}, \mathbf{W}, \mathbf{I}_n)$$

$$\text{(A.86)} \qquad = \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1} + \boldsymbol{\Theta}_t\mathbf{A}_t^{-1}$$

The result then follows from well-known properties of a linear transformation of a matrix-normal random variable, which, in this case, is $\boldsymbol{\Theta}_t$. □

A.5 Densities under $\mathcal{D}_{0:T}$

Claim 5. *With the law of motion*

$$\text{(A.87)} \qquad \mathbf{H}_t = \frac{1}{\beta}h(\mathbf{H}_{t-1})\boldsymbol{\Gamma}_t h(\mathbf{H}_{t-1})' \qquad \boldsymbol{\Gamma}_t \sim B_n(\nu(\beta)/2, 1/2)$$

and $\nu(\beta) = \beta/(1-\beta)$, the density $p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1})$ is given by equation (A.107).

Proof. I derive the density for $\mathbf{H}_t$ as a change of variables from $\boldsymbol{\Gamma}_t$ where, $\mathbf{H}_t = g(\mathbf{H}_{t-1}, \boldsymbol{\Gamma}_t|\boldsymbol{\phi})$ is defined in the statement of the Claim. Rearranging the equation gives

$$g^{-1}(\mathbf{H}_t, \mathbf{H}_{t-1}|\boldsymbol{\phi}) = \beta h(\mathbf{H}_{t-1})^{-1}\mathbf{H}_t h(\mathbf{H}_{t-1})^{-1'} = \boldsymbol{\Gamma}_t \tag{A.88}$$

Hence the density takes the form

$$p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}) = p_{B_n}\left(\beta h(\mathbf{H}_{t-1})^{-1}\mathbf{H}_t h(\mathbf{H}_{t-1})^{-1'}\middle|\nu(\beta)/2, 1/2\right) \cdot |Jg^{-1}| \tag{A.89}$$

where

$$|Jg^{-1}| = |\det(\beta^{1/2}h(\mathbf{H}_{t-1})^{-1})|^{n+1} = \left(\beta^{n/2}|\det(h(\mathbf{H}_{t-1})^{-1}))|\right)^{n+1} \tag{A.90}$$

$$= \beta^{n(n+1)/2}|\det(h(\mathbf{H}_{t-1}))|^{-(n+1)}\,. \tag{A.91}$$

□

Claim 6.

$$p_{\mathcal{D}_t}(\mathbf{B}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_t) = p_N(\mathbf{B}_{t-1}, \mathbf{W}, \mathbf{H}_t^{-1})\,. \tag{A.92}$$

Proof. The claim follows from well-known properties of linear transformations of matrix normal random variables. In this case the random variable being transformed is $\mathbf{V}_t$. □

A.6 Equivalence of densities under $g(\mathbf{S}_t)$ and $\mathcal{D}_{0:T}$

Corollary 3. *Under the definitions of $\mathbf{B}_t \equiv \mathbf{F}_t\mathbf{A}_t^{-1}$ and $\mathbf{H}_t \equiv \mathbf{A}_t\mathbf{A}_t'$ for each t, for any $\mathbf{L}_t, \mathbf{R}_t \in \mathcal{O}_n$*

$$p_{\mathcal{D}_t}(\mathbf{B}_t|\boldsymbol{\phi}, \mathbf{B}_{t-1}, \mathbf{H}_t) = p_{g,\mathcal{S}_t}(\mathbf{F}_t\mathbf{A}_t^{-1}|\mathbf{F}_{t-1}, \mathbf{A}_{t-1}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_t)\,. \tag{A.93}$$

Proof.

$$p_{g,\mathcal{S}_t}(\mathbf{F}_t\mathbf{A}_t^{-1}|\mathbf{F}_{t-1},\mathbf{A}_{t-1},\mathbf{L}_t,\mathbf{R}_t,\mathbf{A}_t) = p_N(\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1},\mathbf{W},(\mathbf{A}_t\mathbf{A}_t')^{-1}) \tag{A.94}$$

$$= p_N(\mathbf{B}_{t-1},\mathbf{W},\mathbf{H}_t^{-1}) \tag{A.95}$$

$$= p_{\mathcal{D}_t}(\mathbf{B}_t|\boldsymbol{\phi},\mathbf{B}_{t-1},\mathbf{H}_t) \tag{A.96}$$

where the first equality holds for any $\mathbf{L}_t, \mathbf{R}_t$ from Claim, 4, the second equality comes from simply substituting from the definitions in the statement of the corollary, and the last equality comes from Claim 6. □

Lemma 6. *For any* $(\mathbf{L}_t, \mathbf{R}_t)$*,*

$$p_{\mathcal{D}_t}(\mathbf{H}_t,\mathbf{B}_t|\boldsymbol{\phi},\mathbf{H}_{t-1},\mathbf{B}_{t-1}) = p(g(\mathbf{A}_t,\mathbf{F}_t)|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t)\,. \tag{A.97}$$

Proof. Factoring the joint densities as

$$\begin{aligned}&p_{\mathcal{D}_t}(\mathbf{H}_t,\mathbf{B}_t|\boldsymbol{\phi},\mathbf{H}_{t-1},\mathbf{B}_{t-1})\\ &\quad= p_{\mathcal{D}_t}(\mathbf{H}_t|\boldsymbol{\phi},\mathbf{H}_{t-1},\mathbf{B}_{t-1})\cdot p_{\mathcal{D}_t}(\mathbf{B}_t|\boldsymbol{\phi},\mathbf{H}_{t-1},\mathbf{B}_{t-1},\mathbf{H}_t)\end{aligned} \tag{A.98}$$

and

$$\begin{aligned}&p(g(\mathbf{A}_t,\mathbf{F}_t)|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t)\\ &\quad= p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t)\cdot p(\mathbf{F}_t\mathbf{A}_t^{-1}|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t,\mathbf{A}_t)\end{aligned} \tag{A.99}$$

The result follows from Lemma 8 and Corollary 3. □

A.7 Additional results on densities of the reduced-form precision matrix

Claim 7. *For* $\mathbf{L}, \widetilde{\mathbf{L}} \in \mathcal{O}_n$*,*

$$p_{B_n}(\beta\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\mathbf{L}_t) = p_{B_n}(\beta\widetilde{\mathbf{L}}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\widetilde{\mathbf{L}}_t)\,.$$

Proof. The result is immediate from Corollary 4.1 in Srivastava (2003). □

Corollary 4. *Under the conditions of Claim 3, if* $\widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, \mathbf{Q}_{t-1} \in \mathcal{O}_n$*, then*

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1}) = p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, \mathbf{A}_{t-1}\mathbf{Q}_{t-1})\,.$$

Proof. Writing out the density explicitly gives

$$\begin{aligned}
&p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, \mathbf{A}_{t-1}\mathbf{Q}_{t-1}) && \text{(A.100)}\\
&\quad= p_{B_n}(\beta\widetilde{\mathbf{L}}_t'(\mathbf{A}_{t-1}\mathbf{Q}_{t-1})^{-1}\mathbf{A}_t\mathbf{A}_t'(\mathbf{A}_{t-1}\mathbf{Q}_{t-1})^{-1'}\widetilde{\mathbf{L}}_t) && \text{(A.101)}\\
&\qquad\cdot \beta^{n(n+1)/2}|\det(\mathbf{A}_{t-1}\mathbf{Q}_{t-1})|^{-(n+1)}\,.\\
&\quad= p_{B_n}(\beta\widetilde{\mathbf{L}}_t'\mathbf{Q}_{t-1}'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1'}\mathbf{Q}_{t-1}\widetilde{\mathbf{L}}_t) && \text{(A.102)}\\
&\qquad\cdot \beta^{n(n+1)/2}(|\det(\mathbf{A}_{t-1})|\underbrace{|\det(\mathbf{Q}_{t-1})|}_{=1})^{-(n+1)}\,.
\end{aligned}$$

Defining $\bar{\mathbf{L}}_t = \mathbf{Q}_{t-1}\widetilde{\mathbf{L}}_t$ gives

$$\begin{aligned}
&p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, \mathbf{A}_{t-1}\mathbf{Q}_{t-1}) && \text{(A.103)}\\
&\quad= p_{B_n}(\beta\bar{\mathbf{L}}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1'}\bar{\mathbf{L}}_t) && \text{(A.104)}\\
&\qquad\cdot \beta^{n(n+1)/2}|\det(\mathbf{A}_{t-1})|^{-(n+1)}\,.
\end{aligned}$$

The result then follows from Claim 7 by noting that, since the orthogonal group is closed under matrix multiplication, $\bar{\mathbf{L}}_t \in \mathcal{O}_n$. □

Corollary 5. *Under the conditions of Claim 3, for any* $\mathbf{Q}_{t-1} \in \mathcal{O}_n$

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1}\mathbf{Q}_{t-1}) = p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{A}_{t-1})\,.$$

where

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{A}_{t-1}) = p_{B_n}(\beta\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1'}) \cdot \beta^{n(n+1)/2}|\det(\mathbf{A}_{t-1})|^{-(n+1)}\,.$$

Proof. The density of $p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{A}_{t-1})$ in equation (A.105) is equivalent to the density $p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, \mathbf{A}_{t-1})$ with $\widetilde{\mathbf{L}}_t = \widetilde{\mathbf{R}}_t = \mathbf{I}_n$. Hence the result follows from Corollary 4. □

Corollary 6. *For* $\widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, \in \mathcal{O}_n$,

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1}) = p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, h(\mathbf{A}_{t-1}\mathbf{A}_{t-1}'))\,.$$

Proof. Note that

$$\mathbf{A}_{t-1}\mathbf{A}_{t-1}' = h(\mathbf{A}_{t-1}\mathbf{A}_{t-1}')h(\mathbf{A}_{t-1}\mathbf{A}_{t-1}')' \tag{A.105}$$

and hence there exists a $\mathbf{P} \in \mathcal{O}_n$ such that $h(\mathbf{A}_{t-1}\mathbf{A}_{t-1}') = \mathbf{A}_{t-1}\mathbf{P}$. The result follows from Corollary 4 by letting $\mathbf{Q}_{t-1} = \mathbf{P}$ and substituting for $h(\mathbf{A}_{t-1}\mathbf{A}_{t-1}')$ with $\mathbf{A}_{t-1}\mathbf{Q}_{t-1}$. □

Corollary 7. *If* $\mathbf{H}_t$ *is defined as* $\mathbf{H}_t \equiv \mathbf{A}_t\mathbf{A}_t'$ *for each t, then*

$$p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1}) = p(\mathbf{H}_t|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, h(\mathbf{H}_{t-1}))\,.$$

Proof. The result is immediate from Corollary 6 by simply making the substitutions defined in the statement of the Corollary. □

Lemma 7.

$$p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}) = p(\mathbf{H}_t|\boldsymbol{\phi}, \widetilde{\mathbf{L}}_t, \widetilde{\mathbf{R}}_t, h(\mathbf{H}_{t-1})) \tag{A.106}$$

where

$$\begin{aligned} p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}) = p_{B_n}&(\beta h(\mathbf{H}_{t-1})^{-1}\mathbf{H}_t h(\mathbf{H}_{t-1})^{-1'}) \\ &\cdot \beta^{n(n+1)/2}|\det(h(\mathbf{H}_{t-1}))|^{-(n+1)}\,. \end{aligned} \tag{A.107}$$

Lemma 8. *Given a parameter point* $\mathbf{A}_{t-1}$ *and defining* $\mathbf{H}_t \equiv \mathbf{A}_t\mathbf{A}_t'$ *for each t, for any* $\mathbf{L}_t, \mathbf{R}_t, \mathbf{Q}_{t-1} \in \mathcal{O}_n$

$$p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}) = p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1}\mathbf{Q}_{t-1})\,. \tag{A.108}$$

Proof. TODO □

In words, the lemma says that the density of $\mathbf{H}_t$ is invariant to the choice of

$\mathbf{L}_t, \mathbf{R}_t \in \mathcal{O}_n$ and depends on $\mathbf{A}_{t-1}$ only through $\mathbf{H}_{t-1} = \mathbf{A}_{t-1}\mathbf{A}'_{t-1}$ and is hence invariant to orthogonal rotations of $\mathbf{A}_{t-1}$ from the right-hand side.

Corollary 8.

$$p(\mathbf{H}_{1:T}|\boldsymbol{\phi}, \mathbf{H}_0) = p(\{\mathbf{A}_t\mathbf{A}'_t\}_{t=1}^T|\boldsymbol{\phi}, \mathbf{L}_{1:T}, \mathbf{R}_{1:T}, \mathbf{A}_0) \tag{A.109}$$

Proof. One can factor the joint densities as

$$p(\mathbf{H}_{1:T}|\boldsymbol{\phi}, \mathbf{H}_0) = \prod_{t=1}^T p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}) \tag{A.110}$$

and

$$p(\{\mathbf{A}_t\mathbf{A}'_t\}_{t=1}^T|\boldsymbol{\phi}, \mathbf{L}_{1:T}, \mathbf{R}_{1:T}, \mathbf{A}_0) = \prod_{t=1}^T p(\mathbf{A}_t\mathbf{A}'_t|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1})\,. \tag{A.111}$$

The result is then immediate from the fact that for each t,

$$p(\mathbf{H}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}) = p(\mathbf{A}_t\mathbf{A}'_t|\boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1})\,, \tag{A.112}$$

which is ensured by Lemma 8. □

A.8 Proof of Theorem 2

Proof of Theorem 2. From the Markovian structure of the laws of motion in equations (26) and (27), one can factor the density of $(\mathbf{H}_{0:T}, \mathbf{B}_{0:T})$ under $\mathcal{D}_{0:T}$ as

$$p_{\mathcal{D}_{0:T}}(\mathbf{H}_{0:T}, \mathbf{B}_{0:T}|\boldsymbol{\phi}) = p_{\mathcal{D}_0}(\mathbf{H}_0, \mathbf{B}_0|\boldsymbol{\phi}) \prod_{t=1}^T p_{\mathcal{D}_t}(\mathbf{H}_t, \mathbf{B}_t|\boldsymbol{\phi}, \mathbf{H}_{t-1}, \mathbf{B}_{t-1}) \tag{A.113}$$

and, similarly, from the laws of motion in equations REF and REF the sequence of random variables $\{g(\mathbf{A}_t, \mathbf{F}_t)\}_{t=0}^{T}$ under $\mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$,

$$\begin{aligned}
& p_{g,\mathcal{S}_{0:T}}(\{g(\mathbf{A}_t, \mathbf{F}_t)\}_{t=1}^{T} | \boldsymbol{\phi}, \mathbf{L}_{1:T}, \mathbf{R}_{1:T}) \\
\text{(A.114)} \qquad & = p_{g,\mathcal{S}_0}(g(\mathbf{A}_0, \mathbf{F}_0) | \boldsymbol{\phi}) \prod_{t=1}^{T} p_{g,\mathcal{S}_t}(g(\mathbf{A}_t, \mathbf{F}_t) | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{L}_t, \mathbf{R}_t)\,.
\end{aligned}$$

The result then follows from the fact that, for each t, by Lemma 6, the density in equation (A.113) equals the density of its time t counterpart in equation (A.114) for any $(\mathbf{L}_t, \mathbf{R}_t)$. □

A.9 Proof of Lemma 1

Lemma 9. $p(\mathbf{H}_{1:T} | \mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T}), \mathbf{H}_0) = p(\mathbf{H}_{1:T} | \mathcal{S}^U_{0:T}(\widetilde{\mathbf{L}}_{1:T}, \widetilde{\mathbf{R}}_{1:T}), \mathbf{H}_0)$

Proof.

$$\begin{aligned}
\text{(A.115)} \quad p(\mathbf{H}_{1:T} | \mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})) &= p(\{\mathbf{A}_t \mathbf{A}'_t\}_{t=1}^{T} | \mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})) \\
\text{(A.116)} \quad &= \prod_{t=1}^{T} p(\mathbf{A}_t \mathbf{A}'_t | \mathcal{S}^U_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T}), \mathbf{A}_{0:t-1}) \\
\text{(A.117)} \quad &= \prod_{t=1}^{T} p(\mathbf{A}_t \mathbf{A}'_t | \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_{t-1})
\end{aligned}$$

TODO now apply the corollary! □

Proof of Lemma 1. For ease of reference I restate the densities given in the main text:

$$\text{(A.118)} \quad p(\mathbf{A}_t \mathbf{A}'_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{L}_t, \mathbf{R}_t) = p_{B_n}(\beta \mathbf{L}'_t \mathbf{A}^{-1}_{t-1} \mathbf{A}_t \mathbf{A}'_t \mathbf{A}^{-1\prime}_{t-1} \mathbf{L}_t) \cdot |\beta^{1/2} \mathbf{L}'_t \mathbf{A}^{-1}_{t-1}|^{n+1}$$

$$\text{(A.119)} \quad p(\mathbf{F}_t \mathbf{A}^{-1}_t | \boldsymbol{\phi}, \mathbf{A}_{t-1}, \mathbf{F}_{t-1}, \mathbf{A}_t \mathbf{A}'_t, \mathbf{L}_t, \mathbf{R}_t) = p_{MN}(\mathbf{F}_{t-1} \mathbf{A}^{-1}_{t-1}, \mathbf{W}, (\mathbf{A}_t \mathbf{A}'_t)^{-1})\,.$$

The argument again requires recognition of the subtlety that changing TODO would, in principle change the density through both the effect of $\widetilde{\mathbf{L}}_t$ on the shock generating $\mathbf{A}_t$ from $\mathbf{A}_{t-1}$, but also through the effect of $\widetilde{\mathbf{L}}_{t-1}$ on the realization of $\mathbf{A}_{t-1}$.

Note that an alternative choice of $\widetilde{\mathbf{L}}_{t-1}$ yields Note that it will suffice to show that

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{L}_t,\mathbf{R}_t) = p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi},\mathbf{A}_{t-1},\widetilde{\mathbf{L}}_t,\widetilde{\mathbf{R}}_t) \tag{A.120}$$

□

A.10 Proof of Lemma ??

***Proof of Lemma* ??.** First, factor the joint density as

(A.121)

$$\begin{aligned} p(\mathbf{H}_t,\mathbf{B}_t|\boldsymbol{\phi},&\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t) \\ &= p(\mathbf{H}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t)\, p(\mathbf{B}_t|\boldsymbol{\phi},\mathbf{A}_{t-1},\mathbf{F}_{t-1},\mathbf{L}_t,\mathbf{R}_t,\mathbf{H}_t) \end{aligned}$$

I prove the property for the two terms of the factorization separately. Simplifying equation (22) gives

$$\mathbf{H}_t = \frac{1}{\beta}\mathbf{A}_{t-1}\mathbf{L}_t h(\boldsymbol{\Gamma}_t)\underbrace{\mathbf{R}_t\mathbf{R}_t'}_{=\mathbf{I}_n} h(\boldsymbol{\Gamma}_t)'\mathbf{L}_t'\mathbf{A}_{t-1}' \tag{A.122}$$

$$= \frac{1}{\beta}\mathbf{A}_{t-1}\mathbf{L}_t \underbrace{h(\boldsymbol{\Gamma}_t)h(\boldsymbol{\Gamma}_t)'}_{\boldsymbol{\Gamma}_t}\mathbf{L}_t'\mathbf{A}_{t-1}' \tag{A.123}$$

$$= \frac{1}{\beta}\mathbf{A}_{t-1}(\mathbf{L}_t\boldsymbol{\Gamma}_t\mathbf{L}_t')\mathbf{A}_{t-1}'\,. \tag{A.124}$$

Now consider alternative orthogonal matrices $(\widetilde{\mathbf{L}}_t,\widetilde{\mathbf{R}}_t)$ substituted into the equation (A.122) in place of $(\mathbf{L}_t,\mathbf{R}_t)$. Identical simplifying steps would ensue and thus the implied law of motion for $\mathbf{H}_t$ is simply $\mathbf{H}_t = (1/\beta)\mathbf{A}_{t-1}(\widetilde{\mathbf{L}}_t\boldsymbol{\Gamma}_t\widetilde{\mathbf{L}}_t')\mathbf{A}_{t-1}'$. Note that $\boldsymbol{\Gamma}_t$ is the only source of randomness in the evolution of $\mathbf{H}_t$. Hence, if it holds that $p(\mathbf{L}_t\boldsymbol{\Gamma}_t\mathbf{L}_t') = p(\widetilde{\mathbf{L}}_t\boldsymbol{\Gamma}_t\widetilde{\mathbf{L}}_t')$, then the result would be proven. By Srivastava (2003)'s Corollary 4.1, it is in fact the case that $p(\mathbf{L}_t\boldsymbol{\Gamma}_t\mathbf{L}_t') = p(\boldsymbol{\Gamma}_t) = p(\widetilde{\mathbf{L}}_t\boldsymbol{\Gamma}_t\widetilde{\mathbf{L}}_t')$ for any $\mathbf{L}_t,\widetilde{\mathbf{L}}_t \in \mathcal{O}_n$. Now consider $\widetilde{\mathbf{A}}_{t-1} = \mathbf{A}_{t-1}\mathbf{Q}_{t-1}$. Substituting into the last expression gives $\mathbf{H}_t = (1/\beta)\mathbf{A}_{t-1}\mathbf{Q}_{t-1}(\widetilde{\mathbf{L}}_t\boldsymbol{\Gamma}_t\widetilde{\mathbf{L}}_t')\mathbf{Q}_{t-1}'\mathbf{A}_{t-1}'$. The product of $\mathbf{Q}_{t-1}\widetilde{\mathbf{L}}_t$ yields an orthogonal matrix, so applying the same result as before proves that the density of $\mathbf{H}_t$ is unchanged.

Next turning to the conditional density of $\mathbf{B}_t$. Defining $\mathbf{V}_t = \mathbf{\Theta}_t(\beta^{-1/2}\mathbf{A}_{t-1}\mathbf{L}_t h(\mathbf{\Gamma}_t))$ well-known properties of the matrix-variate normal distribution,

(A.125)
$$\mathbf{B}_t \sim MN\left(\mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}, \mathbf{W}, (1/\beta)\left([\mathbf{A}_{t-1}\mathbf{L}_t h(\mathbf{\Gamma}_t)\mathbf{R}_t][\mathbf{A}_{t-1}\mathbf{L}_t h(\mathbf{\Gamma}_t)\mathbf{R}_t]'\right)^{-1}\right)$$

□

A.11 Proof of Theorem 2

Proof of Theorem 2. The proof proceeds in two parts. I first show that the density $p(\mathbf{H}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t)$ under the law of the motion defined in equation (22) is identical to the density $p(\mathbf{H}_t|\mathbf{H}_{t-1}, \boldsymbol{\phi})$ under the law of motion

$$\mathbf{H}_t = \frac{1}{\beta}h(\mathbf{H}_{t-1})\mathbf{\Gamma}_t h(\mathbf{H}_{t-1})' \quad \text{for} \quad \mathbf{\Gamma}_t \sim B_n(\nu(\beta)/2, 1/2) \tag{A.126}$$

Indeed, the result is immediate from Lemma **??** by choosing $\widetilde{\mathbf{L}}_t = \mathbf{A}_{t-1}^{-1}h(\mathbf{H}_{t-1})$. That this definition of $\widetilde{\mathbf{L}}_t$ yields an orthogonal matrix is apparent from the fact that $\mathbf{A}_{t-1}$ and $h(\mathbf{H}_{t-1})$ are each "square roots" of $\mathbf{H}_{t-1}$ and hence differ from each other by multiplication by an orthogonal matrix.[22]

One can make a similar statement about the law of motion for $\mathbf{B}_t$ as follows, The density $p(\mathbf{B}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t, \mathbf{A}_t)$ under the law of the motion defined in equation (23) is identical to the density $p(\mathbf{B}_t|\mathbf{B}_{t-1}, \boldsymbol{\phi}, \mathbf{H}_t)$ under the law of motion

$$\mathbf{B}_t = \mathbf{B}_{t-1} + \mathbf{V}_t \quad \text{for} \quad \mathbf{V}_t \sim MN(\mathbf{0}_{m\times n}, \mathbf{W}, \mathbf{H}_t^{-1})\,. \tag{A.127}$$

Substituting into the expression in (**??**) with the definition of $\mathbf{B}_{t-1} = \mathbf{F}_{t-1}\mathbf{A}_{t-1}^{-1}$ and noting that $\mathbf{A}_t^{-1\prime}\mathbf{A}_t^{-1} = (\mathbf{A}_t\mathbf{A}_t')^{-1} = \mathbf{H}_t^{-1}$ immediately gives the representation in (27). □

[22]For example, see Muirhead (1982)'s Theorem Theorem A9.5.

B. Observational Equivalence of Structural and Reduced-Forms

Lemma 10.

$$p_{\mathcal{D}_{0:T}}(\mathbf{y}_{1:T}|\boldsymbol{\phi}) = p_{\mathcal{S}_{0:T}}(\mathbf{y}_{1:T}|\boldsymbol{\phi}) \tag{B.128}$$

Proof. First note that for either $\mathcal{M} \in \{\mathcal{D}_{0:T}, \mathcal{S}^{U}_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})\}$

$$p_{\mathcal{M}}(\mathbf{y}_{1:T}|\boldsymbol{\phi}) = \prod_{t=1}^{T} p_{\mathcal{M}}(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \tag{B.129}$$

where each

$$p_{\mathcal{M}}(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}) = \int_{\mathbf{M}_t \in \mathscr{M}} p(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}, \mathbf{M}_t) p(\mathbf{M}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{M}_t \tag{B.130}$$

$$= \int_{\mathbf{M}_t \in \mathscr{M}} p(\mathbf{y}_t|\mathbf{y}_{t-p:t-1}, \mathbf{M}_t) p(\mathbf{M}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{M}_t \tag{B.131}$$

and

$$p(\mathbf{M}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}) = \int_{\mathbf{M}_{t-1}} p(\mathbf{M}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}, \mathbf{M}_{t-1}) p(\mathbf{M}_{t-1}|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{M}_{t-1} \tag{B.132}$$

$$= \int_{\mathbf{M}_{t-1}} p(\mathbf{M}_t|\boldsymbol{\phi}, \mathbf{M}_{t-1}) p(\mathbf{M}_{t-1}|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{M}_{t-1} \tag{B.133}$$

The line of argument I follow is to show that each $p_{\mathcal{D}_{0:T}}(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}) = p_{\mathcal{S}_{0:T}}(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})$.

One can write each of these densities as a single expression for a given model as

$$p_{\mathcal{M}}(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})$$

$$= \int_{\mathbf{M}_t \in \mathscr{M}} p(\mathbf{y}_t|\mathbf{y}_{t-p:t-1}, \mathbf{M}_t) \tag{B.134}$$

$$\cdot \left(\int_{\mathbf{M}_{t-1}} p(\mathbf{M}_t|\boldsymbol{\phi}, \mathbf{M}_{t-1}) p(\mathbf{M}_{t-1}|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{M}_{t-1} \right) d\mathbf{M}_t \tag{B.135}$$

Consider an alternative parameterization, which one might call a change of variables, defined by $\widetilde{\mathbf{M}}_t = g_t(\mathbf{M}_t)$. Under certain regularity conditions

(B.136)
$$\int_{\mathbf{M}_{t-1} \in \mathscr{M}_{t-1}} p(\mathbf{M}_t | \boldsymbol{\phi}, \mathbf{M}_{t-1}) p(\mathbf{M}_{t-1} | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \, d\mathbf{M}_{t-1}$$
(B.137)
$$= \int_{\mathbf{M}_{t-1} \in \mathscr{M}_{t-1}} p(\mathbf{M}_t | \boldsymbol{\phi}, g(\mathbf{M}_{t-1})) p(g(\mathbf{M}_{t-1}) | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) | \det(dg(\mathbf{M})) | \, d\mathbf{M}_{t-1}$$
(B.138)
$$= \int_{\widetilde{\mathbf{M}}_{t-1} \in g(\mathscr{M}_{t-1})} p(\mathbf{M}_t | \boldsymbol{\phi}, \widetilde{\mathbf{M}}_{t-1}) p(\widetilde{\mathbf{M}}_{t-1} | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) d\widetilde{\mathbf{M}}_{t-1}$$

Beginning with $t = 0$, consider

$$p_{\mathcal{S}_{0:T}}(\mathbf{y}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1})$$
(B.139)
$$= \int_{\mathbf{S}_t \in \mathcal{S}_t} p(\mathbf{y}_t | \mathbf{y}_{t-p:t-1}, \mathbf{S}_t)$$
(B.140)
$$\cdot \left(\int_{\mathbf{S}_{t-1} \in \mathcal{S}_{t-1}} p(\mathbf{S}_t | \boldsymbol{\phi}, \mathbf{S}_{t-1}) p(\mathbf{S}_{t-1} | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \, d\mathbf{S}_{t-1} \right) d\mathbf{S}_t$$

Where $p(\mathbf{S}_t | \boldsymbol{\phi}, \mathbf{S}_{t-1})$ is derived in Now consider the change of variables where $\mathbf{D}_t = g_t(\mathbf{S}_t)$

If each $g_t(\cdot)$ is injective, differentiable, with continuous partial derivatives then for each t

$$\int_{\mathbf{S}_t} p(\mathbf{y}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}, \mathbf{S}_t) p(\mathbf{S}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \, d\mathbf{S}_t$$
(B.141)
$$= \int_{g(\mathbf{S}_t)} p(\mathbf{y}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}, g(\mathbf{S}_t)) p(g(\mathbf{S}_t) | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \, | \det(Dg(\mathbf{S}_t)) | d\mathbf{S}_t$$
(B.142)
$$= \int_{\mathbf{D}_t} p(\mathbf{y}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}, \mathbf{D}_t) p(\mathbf{D}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \, d\mathbf{D}_t$$

defining $g_t(\mathbf{S}_t) = \mathbf{D}_t$, then $p(\mathbf{D}_t | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) = p(g(\mathbf{S}_t) | \boldsymbol{\phi}, \mathbf{y}_{1:t-1}) \, | \det(Dg(\mathbf{S}_t)) |$.

hence the equivalence of each term in

$$stuff \tag{B.143}$$

then the result follows. Where

$$p(\mathbf{S}_t|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}) = \int_{\mathbf{S}_{t-1}} p(\mathbf{S}_t|\boldsymbol{\phi}, \mathbf{S}_{t-1}) p(\mathbf{S}_{t-1}|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{S}_{t-1}\,. \tag{B.144}$$

and hence

$$p(g(\mathbf{S}_t)|\boldsymbol{\phi}, \mathbf{y}_{1:t-1}) = \int_{\mathbf{S}_{t-1}} p(g(\mathbf{S}_t)|\boldsymbol{\phi}, \mathbf{S}_{t-1}) p(\mathbf{S}_{t-1}|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, d\mathbf{S}_{t-1}\,. \tag{B.145}$$

$$= \int_{g(\mathbf{S}_{t-1})} p(g(\mathbf{S}_t)|\boldsymbol{\phi}, g(\mathbf{S}_{t-1})) p(g(\mathbf{S}_{t-1})|\boldsymbol{\phi}, \mathbf{y}_{1:t-1})\, |\det(Dg(\mathbf{S}_{t-1}))| d\mathbf{S}_{t-1}\,. \tag{B.146}$$

□

Theorem 3. *Let $\mathcal{D}_{0:T}$ and $\mathcal{S}^{U}_{0:T}(\mathbf{L}_{1:T}, \mathbf{R}_{1:T})$ be as defined in the main text and let $\mathbf{D}_t = g(\mathbf{S}_t)$, $g^{-1}(\mathbf{D}_t) = \mathbf{S}_t$. If*

1. $p_{\mathcal{S}_{0:T}}(\boldsymbol{\phi}) = p_{\mathcal{D}_{0:T}}(\boldsymbol{\phi}) = p(\boldsymbol{\phi})$
2. $p(\mathbf{S}_0|\boldsymbol{\phi})$ *is such that*

$$p(g(\mathbf{S}_0)|\boldsymbol{\phi})) = p(\mathbf{D}_0|\boldsymbol{\phi}) = p_{NW}(\mathbf{D}_0|\overline{\mathbf{B}}_{0|0}, \mathbf{C}_{0|0}, d_{0|0}, \boldsymbol{\Psi}^{-1}_{0|0})$$

then for any $\mathbf{y}_{1:T}$

$$p_{\mathcal{D}_{0:T}}(\mathbf{y}_{1:T}) = p_{\mathcal{S}_{0:T}}(\mathbf{y}_{1:T})\,. \tag{B.147}$$

Proof. The proof is nearly immediate from Lemma 10.

$$p_{D_{0:T}}(\mathbf{y}_{1:T}) = \int_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} p_{D_{0:T}}(\mathbf{y}_{1:T}|\boldsymbol{\phi})p_{D_{0:T}}(\boldsymbol{\phi})d\boldsymbol{\phi} \tag{B.148}$$

$$= \int_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} p_{S_{0:T}}(\mathbf{y}_{1:T}|\boldsymbol{\phi})p_{S_{0:T}}(\boldsymbol{\phi})d\boldsymbol{\phi} \tag{B.149}$$

$$= p_{S_{0:T}}(\mathbf{y}_{1:T}) \tag{B.150}$$

Where the second equality follows from the premise of the theorem and the equivalence of each data-density term from Lemma 10. □

C. Gibbs Sampler for Reduced-Form Parameters

This 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 $\mathbf{D}_{0:T}$ backwards conditional on $\boldsymbol{\phi}^*$.

C.1 Block 1: $\mathbf{W}|\mathbf{y}_{1:T}, \beta, D_{0:T}$

Given a draw of the history of latent states $\mathbf{D}_{0:T}$, the matrix of shocks $\mathbf{V}_t$ to the linear coefficients in (23) become observable via

$$\mathbf{V}_t = \mathbf{B}_t - \mathbf{B}_{t-1}\ . \tag{C.151}$$

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

$$p(\mathbf{W}|Y, \beta, \mathbf{D}_{0:T}) \sim IW(\Psi_{0:T}, \nu_{0:T}) \tag{C.152}$$

where

$$\Psi_{0:T} = \Psi_0 + \Psi_{1:T} \tag{C.153}$$

$$\nu_{0:T} = \nu_0 + \nu_{1:T} \tag{C.154}$$

and

$$\Psi_{1:T} = \sum_{t=1}^{T} \mathbf{V}_t \mathbf{H}_t \mathbf{V}_t' \tag{C.155}$$

$$\nu_{1:T} = \sum_{t=1}^{T} n = Tn \,. \tag{C.156}$$

Appendix E gives a derivation of these expressions.

C.2 Block 2: $\beta, \mathbf{D}_{0:T} | \mathbf{y}_{1:T}, \mathbf{W}$

Sampling from Block 2 entails sampling from the joint distribution of $\beta, \mathbf{D}_{0:T} | \mathbf{y}_{1:T}, \mathbf{W}$. I accomplish this by first sampling from the marginal distribution of $\beta | \mathbf{y}_{1:T}, \mathbf{W}$ and subsequently drawing from the conditional distribution of $\mathbf{D}_{0:T} | \mathbf{y}_{1:T}, \mathbf{W}, \beta$.

C.2.1 Step 2a: $\beta | \mathbf{y}_{1:T}, \mathbf{W}$

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

$$\alpha\left(\beta^* | \mathbf{y}_{1:T}, \mathbf{W}\right) = \min\left\{ \frac{p\left(\beta^*, \mathbf{W}^{(i)} | \mathbf{y}_{1:T}\right) q(\beta^{(i-1)} | \beta^*)}{p\left(\beta^{(i-1)}, \mathbf{W}^{(i)} | \mathbf{y}_{1:T}\right) q(\beta^* | \beta^{(i-1)})},\ 1 \right\} \,. \tag{C.157}$$

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 (C.157) cancels. Let

$$k\left(\mathbf{W}, \beta | \mathbf{y}_{1:T}\right) = p(\mathbf{W}, \beta) p(\mathbf{y}_{1:T} | \beta, \mathbf{W}) \,, \tag{C.158}$$

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 (C.157). Hence, we can calculate α as

$$\alpha\left(\beta^*|\mathbf{y}_{1:T},\mathbf{W}\right)=\min\left\{\frac{k\left(\mathbf{W}^{(i)},\beta^*|\mathbf{y}_{1:T}\right)}{k\left(\mathbf{W}^{(i)},\beta^{(i-1)}|\mathbf{y}_{1:T}\right)},\ 1\right\}\ , \tag{C.159}$$

so long as we can calculate $k(\mathbf{W},\beta|\mathbf{y}_{1:T})$ pointwise. One can indeed evaluate $p(\mathbf{y}_{1:T}|\beta,\mathbf{W})$ pointwise by using the recursive filtering algorithm described in Prado and West (2010). For completeness, and because of its lesser familiarity in economics, I summarize the steps of the recursive filter and its use for likelihood evaluation in Table A-1.

Write the joint density of $\mathbf{y}_t$ and $\mathbf{D}_t$ as

$$p(\mathbf{y}_t,\mathbf{D}_t|\boldsymbol{\phi},\mathbf{y}_{1:t-1})=p(\mathbf{y}_t|\boldsymbol{\phi},\mathbf{y}_{1:t-1},\mathbf{D}_t)\ p(\mathbf{D}_t|\boldsymbol{\phi},\mathbf{y}_{1:t-1})\ . \tag{C.160}$$

Let $\Phi=(M,C,d,D)$ denote the sufficient statistics for the $\mathcal{NW}$ distribution. We can evaluate the likelihood using the following recursions.

Step 0: Posterior for dynamic parameters from $t-1$. Suppose the posterior for $S_{t-1}|\theta,Y_{0:t-1}$ is of the form $\mathcal{NW}(\Phi_{t-1|t-1})$, where the elements of $\Phi_{t-1|t-1}$ are specified in Step 0 of Table A-1.

Step 1: Prior for dynamic parameters entering t. The functional forms for the state transition equations and their innovations yield a prior for $S_t|\theta,Y_{0:t-1}$ of the form $\mathcal{NW}(\Phi_{t|t-1})$, where the elements of $\Phi_{t|t-1}$ are specified in Step 1 of Table A-1, see Chapter 10 of Prado and West (2010).

Step 1.5: Evaluate the forecast density.

Step 2: Posterior for dynamic parameters. Update beliefs about (S_t) according to Bayes rule,

$$p(S_t|\theta,Y_{0:t})=\frac{p(y_t|\theta,Y_{0:t-1},S_t)p(S_t|\theta,Y_{0:t-1})}{p(y_t|\theta,Y_{0:t-1})}\ , \tag{C.161}$$

where the posterior for (H_t, B_t) is summarized by the sufficient statistics $\Phi_{t|t}$ that can be computed according to the expressions in Step 2 of Table A-1.

Step 3: Return to Step 1 for $t+1$. In the words of Uhlig (1994), "the game can begin anew."

Following Step 2, one can evaluate the likelihood

$$p(y_t|\theta,Y_{1:t-1}) = \int_{S_t} p(y_t|\theta, Y_{1:t-1}, S_t)\, p(S_t|\theta, Y_{1:t-1}) dS_t \ , \tag{C.162}$$

using the familiar closed-form expression for the marginal likelihood of a VAR, where the VAR here happens to have a single observation, y_t.

(C.163) MDD EQUATION HERE

Iterating on these steps until $t = T$ is reached, one can evaluate the likelihood for the full sequence of observations as the product of conditional likelihoods

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|\theta, Y_{t-1}) \tag{C.164}$$

where each $p(y_t|\theta, Y_{0:t-1})$ is evaluated via (C.163).

C.2.2 Step 2b: $\mathbf{D}_{0:T}|\mathbf{y}_{1:T}, \beta, \mathbf{W}$

The final step in the MCMC algorithm is a sweep through the recursive backwards "smoothing" algorithm of Prado and West (2010) for $\mathbf{D}_{0:T}$.[23] For completeness I summarize the smoother in Table A-2. The draw proceeds backwards from the end the forward filtering algorithm used to evaluate the likelihood in Step 2a.

D. Computational considerations for the MCMC Algorithm

A few aspects of an efficient implementation of the MCMC algorithm are worth calling attention to. First, The smoothing algorithm used in Step 2b makes

[23] See section 10.4.5 of Prado and West (2010).

use of some objects computed during a run of the forward filtering algorithm. In particular, the terminal values of $(h_T, D_T^{-1}, M_T, C_T)$ summarize the posterior distribution of (H_T, B_T) from which the backwards smoother is initialized, and the entire sequences of $a_{1:T}$, $R_{1:T}$, $C_{1:T}$, and $M_{1:T}$ are used in the smoother as well. Since the filtering step would have been run to evaluate the likelihood function during the MH algorithm in Step 2a, the values of the state variables can be stored and then reused in step 2b.

Second, the posterior smoothing algorithm in Step 2b requires sampling repeatedly from a potentially high-dimensional multivariate normal distribution, whose covariance matrix is given by $V_{t|t+1} = H_{t+1}^{-1} \otimes C_t^*$. The kronecker structure of the covariance means that the distribution takes the special form known as matrix normal distribution, which allows for dramatically faster sampling than naively computing $V_{t|t+1}$ and drawing from the multivariate normal with $V_{t|t+1}$ as its covariance. Namely a random draw of b_t can be obtained via a random draw of a matrix $\mathbf{X}_{m\times n}$ populated with iid univariate normal random draws and then setting

$$b_t = \mu_{t|t+1} + \text{vec}\big(\mathcal{U}(C_t^*)' \, \mathbf{X} \, \mathcal{U}(H_{t+1}^{-1})\big) \tag{D.165}$$

where $\mathcal{U}(.)$ denotes the upper-triangular Cholesky decomposition. This method of sampling generates a speed gain of more than an order of magnitude in `MATLAB` in the context of the six variable application.

E. Derivation of Block 1 in the MCMC Algorithm

One can derive the expressions for Block 1 of the MCMC algorithm as follows. Conditional on β and the dynamic latent variables $\mathbf{D}_{t-1:t}$, the "likelihood" for $\mathbf{W}$ is

$$p(\mathbf{V}_t|\mathbf{W},\mathbf{D}_{t-1:t}, \beta, \mathbf{y}_{1:T}) = p_{MN}(\mathbf{0}_{m\times n}, \mathbf{W}, \mathbf{H}_t^{-1}) \tag{E.166}$$

$$= (2\pi)^{-mn/2}|\mathbf{H}_t^{-1}|^{-m/2} \, |\mathbf{W}|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\mathbf{H}_t\mathbf{V}_t'\mathbf{W}^{-1}\mathbf{V}_t\right]\right\} \tag{E.167}$$

Letting $c_{\mathbf{W},t} = (2\pi)^{-mn/2}|\mathbf{H}_t^{-1}|^{-m/2}$, and using the trace operator's invariance to cyclic permutations, one can write

$$p(\mathbf{V}_t|\mathbf{W}, \mathbf{D}_{t-1:t}, \beta, \mathbf{y}_{1:T}) = p(\mathbf{V}_t|\mathbf{W}, \mathbf{D}_{t-1:t}) \tag{E.168}$$

$$= c_{\mathbf{W},t}\, |\mathbf{W}|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\mathbf{V}_t\mathbf{H}_t\mathbf{V}_t'\mathbf{W}^{-1}\right]\right\} \tag{E.169}$$

and, letting $\Psi_t = \mathbf{V}_t\mathbf{H}_t\mathbf{V}_t'$, even more compactly as

$$p(\mathbf{V}_t|\mathbf{W}, \mathbf{D}_{t-1:t}, \beta, \mathbf{y}_{1:T}) = c_{\mathbf{W},t}\, |\mathbf{W}|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\Psi_t\mathbf{W}^{-1}\right]\right\}. \tag{E.170}$$

The likelihood for $\mathbf{W}$ given the full sequence $\mathbf{V}_{1:T}$ is

$$p(\mathbf{V}_{1:T}|\mathbf{W}, \mathbf{D}_{0:T}) = \prod_{t=1}^{T} c_{\mathbf{W},t}\, p(\mathbf{V}_t|\mathbf{W}, \mathbf{D}_{t-1:t}). \tag{E.171}$$

Defining $c_{\mathbf{W}} = \prod_{t=1}^{T} c_{\mathbf{W},t}$ gives

$$p(\mathbf{V}_{1:T}|\mathbf{W}, \mathbf{D}_{0:T})$$

$$= c_{\mathbf{W}}\, |\mathbf{W}|^{-Tn/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\text{tr}\left[\Psi_t\mathbf{W}^{-1}\right]\right\} \tag{E.172}$$

$$= c_{\mathbf{W}}\, |\mathbf{W}|^{-Tn/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{t=1}^{T}\Psi_t\mathbf{W}^{-1}\right]\right\} \tag{E.173}$$

$$= c_{\mathbf{W}}\, |\mathbf{W}|^{-Tn/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\left(\sum_{t=1}^{T}\Psi_t\right)\mathbf{W}^{-1}\right]\right\} \tag{E.174}$$

Letting $\Psi_{1:T} = \sum_{t=1}^{T}\Psi_t$ one can write

$$p(\mathbf{V}_{1:T}|\mathbf{W}, \mathbf{D}_{0:T}) = c_{\mathbf{W}}\, |\mathbf{W}|^{-Tn/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\Psi_{1:T}\mathbf{W}^{-1}\right]\right\}. \tag{E.175}$$

If the prior for $\mathbf{W}$ is $IW(\Psi_0, \nu_0)$ with density

$$p(\mathbf{W}) = c_{IW}\, |\mathbf{W}|^{-(\nu_0+m+1)/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\Psi_0\mathbf{W}^{-1}\right]\right\}, \tag{E.176}$$

then the conditional posterior takes the form

$$p(\mathbf{W}|\mathbf{D}_{0:T}) = c_{IW} c_{\mathbf{W}} \left|\mathbf{W}\right|^{-((v_0+Tn)+m+1)/2} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\left(\Psi_0 + \Psi_{1:T}\right)\mathbf{W}^{-1}\right]\right\}. \tag{E.177}$$

F. Proofs of Claims

Proof of Claim 1. I derive the density of $\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}, \mathbf{L}_t, \mathbf{R}_t$ as a change of variables from $\boldsymbol{\Gamma}_t$.

$$\mathbf{A}_t = g(h(\boldsymbol{\Gamma}_t)|\mathbf{A}_{t-1}, \boldsymbol{\phi}) = \beta^{-1/2}\mathbf{A}_{t-1}\mathbf{L}_t h(\boldsymbol{\Gamma}_t)\mathbf{R}_t \tag{F.178}$$

$$h(\boldsymbol{\Gamma}_t) = g^{-1}(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}) = \sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{R}_t' \tag{F.179}$$

and

$$\boldsymbol{\Gamma}_t = h^{-1}(g^{-1}(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi})) \tag{F.180}$$

where $h^{-1}(\boldsymbol{\Omega}) = \boldsymbol{\Omega}\boldsymbol{\Omega}'$. Hence the density of $\mathbf{A}_t$ has the form

$$p(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}) = p_{\boldsymbol{\Gamma}}\left(h^{-1}(g^{-1}(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}))\right) \cdot |J(h^{-1})| \cdot |J(g^{-1})|. \tag{F.181}$$

Filling in the exact expressions for the terms in equation (F.181), the first term is the density of multivariate beta distribution, which I denote

$$p_{\boldsymbol{\Gamma}}\left(h^{-1}(g^{-1}(\mathbf{A}_t|\mathbf{A}_{t-1}, \boldsymbol{\phi}))\right) = p_{B_n}(\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\underbrace{\mathbf{R}_t'\mathbf{R}_t}_{\mathbf{I}_n}\mathbf{A}_t'\mathbf{A}_{t-1}^{-1\prime}\mathbf{L}_t\sqrt{\beta}). \tag{F.182}$$

The term $|J(h^{-1})|$ is the Jacobian associated with the Cholesky factor of the random matrix $\boldsymbol{\Gamma}_t$, which is known to take the form

$$|J(h^{-1})| = \prod_{i=1}^{n}[h(\boldsymbol{\Gamma}_t)]_{ii}^{n-i} = \prod_{i=1}^{n}\left[\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{R}_t'\right]_{ii}^{n-i}. \tag{F.183}$$

(Alternatively, one might derive the elements of the density coming from the first two terms in equation (F.181) immediately from Srivastava (2003)'s Corollary 3.1, which would give identical expressions).

The term $|J(g^{-1})|$ is the Jacobian associated with the linear transformation from $h(\mathbf{\Gamma}_t)$ to $\mathbf{A}_t$ and is thus given by

$$|J(g^{-1})| = \det\left(\mathbf{R}_t' \otimes (\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1})\right) \tag{F.184}$$

$$= \det(\mathbf{R}_t')^n \cdot \det(\sqrt{\beta}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1})^n \tag{F.185}$$

$$= \beta^{n/2} \underbrace{\det(\mathbf{R}_t)^n}_{=1} \underbrace{\det(\mathbf{L}_t)^n}_{=1} \det(\mathbf{A}_{t-1}^{-1})^n \tag{F.186}$$

$$= \beta^{n/2} \det(\mathbf{A}_{t-1}^{-1})^n \tag{F.187}$$

$\square$

Proof of Claim 3. Recall the definition

$$\mathbf{A}_t\mathbf{A}_t' = g_{\mathbf{H}}(\mathbf{\Gamma}_t, \mathbf{S}_{t-1}) = \frac{1}{\beta}\left(\mathbf{A}_{t-1}\mathbf{L}_t h(\mathbf{\Gamma}_t)\mathbf{R}_t\right)\left(\mathbf{R}_t' h(\mathbf{\Gamma}_t)'\mathbf{L}_t'\mathbf{A}_{t-1}'\right) \tag{F.188}$$

Rearranging this definition, one can write

$$\beta\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1'}\mathbf{L}_t = h(\mathbf{\Gamma}_t)\underbrace{\mathbf{R}_t\mathbf{R}_t'}_{=\mathbf{I}_n} h(\mathbf{\Gamma}_t)' = h(\mathbf{\Gamma}_t)h(\mathbf{\Gamma}_t)' = \mathbf{\Gamma}_t \tag{F.189}$$

hence the inverse transformation is given by

$$g_{\mathbf{H}}^{-1}(\mathbf{A}_t\mathbf{A}_t', \mathbf{S}_{t-1}) = (\beta^{1/2}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1})(\mathbf{A}_t\mathbf{A}_t')(\beta^{1/2}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1})' = \mathbf{\Gamma}_t \tag{F.190}$$

From the definition of $\mathbf{\Gamma}_t$ the density of $\mathbf{A}_t\mathbf{A}_t'$ then has the form

$$p(\mathbf{A}_t\mathbf{A}_t'|\boldsymbol{\phi}, \mathbf{A}_{t-1}) = p_{B_n}(\beta\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}\mathbf{A}_t\mathbf{A}_t'\mathbf{A}_{t-1}^{-1'}\mathbf{L}_t) \cdot |J g_{\mathbf{H}_t}^{-1}|\,. \tag{F.191}$$

From the symmetry of $\mathbf{A}_t\mathbf{A}_t'$, see Table 6.1 row (i) in Magnus and Neudecker (1980),

$$
\begin{aligned}
&\text{(F.192)} \qquad |J(\mathbf{A}_t\mathbf{A}_t', \boldsymbol{\Gamma}_t|\mathbf{S}_{t-1})| = |J g_{\mathbf{H}_t}^{-1}| = |\beta^{1/2}\mathbf{L}_t'\mathbf{A}_{t-1}^{-1}|^{n+1} \\
&\text{(F.193)} \qquad = \left(\beta^{n/2} \underbrace{|\det(\mathbf{L}_t')|}_{=1} |\det(\mathbf{A}_{t-1}^{-1})|\right)^{n+1} \\
&\text{(F.194)} \qquad = \beta^{n(n+1)/2} |\det(\mathbf{A}_{t-1}^{-1})|^{n+1} \\
&\text{(F.195)} \qquad = \beta^{n(n+1)/2} |\det(\mathbf{A}_{t-1})|^{-(n+1)}\,.
\end{aligned}
$$

□

TABLE A-1

RECURSIVE FILTER FOR LIKELIHOOD EVALUATION MARGINAL OF DYNAMIC PARAMETERS

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

Notes: The table summarizes results given in Prado and West (2010).

TABLE A-2

BACKWARDS SIMULATION SMOOTHER FOR DYNAMIC PARAMETERS

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

References

AMIR-AHMADI, P. AND T. DRAUTZBURG (2017): "Identification Through Heterogeneity," *Federal Reserve Bank of Philadelphia, working paper*, 17-11.

ANTOLÍN-DÍAZ, J. AND J. F. RUBIO-RAMÍREZ (2017): "Narrative Sign Restrictions," *Mimeo, Emory University*.

AUERBACH, A. AND Y. GORODNICHENKO (2012): "Fiscal Multipliers in Recession and Expansion," in *Fiscal Policy after the Financial Crisis*, ed. by A. Alesina and F. Giavazzi, National Bureau of Economic Research, University of Chicago Press, 63–98.

BAUMEISTER, C. AND J. D. HAMILTON (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 83, 1963–1999.

BAUMEISTER, C. AND G. PEERSMAN (2013a): "The role of time-varying price elasticities in accounting for volatility changes in the crude oil market," *Journal of Applied Econometrics*, 28, 1087–1109.

——— (2013b): "Time-Varying Effects of Oil Supply Shocks on the US Economy," *American Economic Journal: Macroeconomics*, 5, 1–28.

CANOVA, F. AND G. DE NICOLO (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

CANOVA, F. AND L. GAMBETTI (2009): "Structural changes in the US economy: Is there a role for monetary policy?" *Journal of Economic Dynamics and Control*, 33, 477–490.

CLARK, T. E. AND F. RAVAZZOLO (2015): "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics*, 30, 551–575.

COGLEY, T. AND T. J. SARGENT (2001): "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, 16, 331–373.

——— (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262 – 302.

DOAN, T., R. LITTERMAN, AND C. SIMS (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3, 1–100.

FOX, E. B. AND M. WEST (2014): "Autoregressive models for variance matrices: Stationary inverse Wishart processes," ArXiv:1107.5239.

HOFMANN, B., G. PEERSMAN, AND R. STRAUB (2012): "Time variation in U.S. wage dynamics," *Journal of Monetary Economics*, 59, 769–783.

KOOP, G. AND D. KOROBILIS (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177, 185–198.

MAGNUS, J. R. AND H. NEUDECKER (1980): "The Elimination Matrix: Some Lemmas and Applications," *SIAM Journal on Algebraic Discrete Methods*, 1, 422–449.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc.

PRADO, R. AND M. WEST (2010): *Time Series: Modeling, Computation, and Inference*, Chapman & Hall/CRC.

PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821–852.

QUINTANA, J. M. AND M. WEST (1987): "An analysis of international exchange rates using multivariate DLM's," *The Statistician*, 275–281.

RAMEY, V. AND S. ZUBAIRY (2017): "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data," *Journal of Political Economy*, Forthcoming.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, 77, 665–696.

Sims, C. A. (1993): "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators and Forecasting*, University of Chicago Press, 179–212.

Sims, C. A., D. F. Waggoner, and T. Zha (2008): "Methods for inference in large multiple-equation Markov-switching models," *Journal of Econometrics*, 146, 255 – 274.

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

Srivastava, M. (2003): "Singular Wishart and multivariate beta distributions," *Ann. Statist.*, 31, 1537–1560.

Uhlig, H. (1994): "On Singular Wishart and Singular Multivariate Beta Distributions," *The Annals of Statistics*, 22, 395–405.

——— (1997): "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica*, 65, pp. 59–73.

——— (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381 – 419.