Bayesian Estimation of Fractionally Integrated Vector Autoregressions and an Application to Identified Technology Shocks

Ross Doppelt and Keith O'Hara*

May 14, 2018

Abstract

We introduce a new method for Bayesian estimation of fractionally integrated vector autoregressions (FIVARs). The FIVAR, which nests a standard VAR as a special case, allows each series to exhibit long memory, meaning that low frequencies can play a dominant role — a salient feature of many macroeconomic and financial time series. Although the parameter space is typically high-dimensional, our inferential procedure is computationally tractable and relatively easy to implement. We apply our methodology to the identification of technology shocks, an empirical problem in which business-cycle predictions depend on carefully accounting for low-frequency fluctuations.

*Doppelt: Penn State and Carnegie Mellon. O'Hara: New York University. Contact: rdoppelt@andrew.cmu.edu and keith.ohara@nyu.edu. We thank Jeff Campbell, David Childers, Fallaw Sowell, and various seminar participants for helpful discussions. Any errors are our own.

1 Introduction

Bayesian vector autoregressions (VARs) are common tools for empirical macroeconomics. As surveyed by Del Negro and Schorfheide (2011) and Koop and Korobilis (2010), the literature on Bayesian VARs has produced valuable methods for describing data, producing forecasts, and identifying structural shocks.[1] We build on this body of work by providing a computationally tractable method for Bayesian estimation of fractionally integrated VARs (FIVARs). Traditionally, VAR practitioners allow variables to enter in levels or differences. In a FIVAR, the j^{th} variable in the VAR has been differenced δ_j times — where δ_j need not be an integer — and these differencing parameters are estimated simultaneously with the VAR parameters, rather than imposed ex ante. This parsimonious modification of a standard VAR can have a substantial effect on the dynamics of the system. Fractionally integrated series can exhibit long memory, meaning that the correlation between a variable and its own lags decays extremely slowly relative to a VAR. In practice, this statistical property of the model can have substantive implications for the identification of economic shocks.

Why generalize the VAR to include fractional integration, and why take a Bayesian approach to doing so? A number of econometricians have argued for the relevance of long-memory models. Henry and Zaffaroni (2003) survey the literature and contend that the long-memory paradigm provides a natural framework for approaching both macro and financial time series. In finance, where many series are observed on a daily basis, fractionally integrated models have proven their utility.[2] In macro, where time series are shorter, we need to investigate the data more carefully. Although many macroeconomic variables appear persistent, sample size is a legitimate concern: Accumulating information about low-frequency oscillations is inherently slow, so frequentist estimators may perform poorly with only 60 years of data. This fact has lead some macroeconometricans to view fractional integration with skepticism.[3] We're going to take the fractional-integration hypothesis seriously, but from a Bayesian perspective. With Bayesian VARs, it's common to fit densely parameterized models, coupled with an informative prior that shrinks most of the coefficients toward zero. That way, the model entertains the possibility of correlations across many variables at many lags, but only if the likelihood contains enough evidence to overcome the prior. We approach FIVARs in

[1] Recent discussions of forecasting with Bayesian VARs can be found in Giannone et al. (2015), Koop (2013), and Bańbura et al. (2010).

[2] For example, in their seminal studies of realized volatility, Andersen et al. (2001, 2003) argue that a Gaussian FIVAR fits their data well. Working with univariate models, Bhardwaj and Swanson (2006) find that fractionally integrated ARIMA models do a good job of forecasting many financial time series. See Baillie (1996) for a review of earlier work.

[3] One prominent skeptic was Granger (1997): "The $I(1)$ assumption can be replaced by $I(d)$ where d is a fraction, so that some of the variables are fractionally integrated. Outside of some special cases in finance, I do not think that this is an especially plausible class of models to consider and am not convinced that they have been widely discovered in practice in macrodata." (p. 173) In his conclusion, though, he remarks: "Use of non-numerical information, such as personal beliefs or experiences as suggested by the Bayesians should be considered, has also been known to produce some disagreement. I personally lack sufficient self confidence to be a formal Bayesian, stating a specific prior, but am happy that not everyone has this personal characteristic." (p. 175)

the same spirit. Shrinking the differencing parameters toward zero provides a clean and transparent way of encoding skepticism toward fractional integration, while still allowing the data to update our beliefs. Our priors are centered on a conventional VAR, but in our applied work, our posteriors indicate that fractional integration helps account for low-frequency variation in key macroeconomic variables, such as inflation and interest rates.

Methodologically, our main contribution is to provide the first computationally viable algorithm for Bayesian estimation of FIVARs. Frequentist estimates of FIVARs are often based on maximum likelihood or, more commonly, quasi-maximum likelihood.[4] In the Bayesian literature, several authors have proposed sampling algorithms for univariate ARIMA(p, d, q) processes with non-integer values of d.[5] Unfortunately, extending these methods to multivariate systems is not straightforward. Although Ravishanker and Ray (1997, 2002) claim to provide a posterior sampler for FIVAR processes, their approach actually produces estimates for a different class of fractionally integrated models, with different low-frequency properties.[6] In addition, they use a variant of the random-walk Metropolis-Hastings algorithm, and posterior sampling will typically be more efficient if any of the conditional distributions have a conjugate (or approximately conjugate) form.

Given the growing popularity of Bayesian methods, we suspect that computational barriers are a major reason why empirical macroeconomists have been slow to adopt fractionally integrated models. In principle, one could use a generic sampling algorithm, such as Metropolis-Hastings, to draw from the posterior. However, Bayesian estimation of VARs (and, by extension, FIVARs) usually involves a large number of parameters, so it's advantageous to exploit the structure of the model to construct an efficient sampler. Using frequency-domain methods, we begin by formulating an approximation to the posterior that has a semi-conjugate form: Conditional on the differencing parameters, the autoregressive coefficients and innovation variances have a normal-Wishart structure. Hence, we can marginalize over the VAR parameters and sample from the (low-dimensional) distribution of differencing parameters. Then, given the differencing parameters, we can draw the VAR parameters directly from the normal-Wishart distribution. Using a sequential importance sampler, along the lines of Herbst and Schorfheide (2014, 2015), we also discuss how to reweight draws from the approximate posterior to perform exact posterior inference. It turns out that the approximate posterior, besides being easy to implement, is often close to the exact posterior. This fact greatly reduces the computational burden of exact posterior inference.

[4] Sowell (1989b) and Sela and Hurvich (2009) discuss exact Gaussian maximum likelihood. Hosoya (1996, 1997) provides the theory for quasi-maximum likelihood based on Whittle's frequency-domain approximation.

[5] Examples include Koop et al. (1997), Pai and Ravishanker (1998), Hsu and Breidt (2003), and Graves et al. (2015).

[6] We discuss the issues with Ravishanker and Ray's method in a separate note; see Doppelt and O'Hara (2018). In general, the class of models for which the Ravishanker-Ray algorithm produces posterior draws will not coincide with a FIVAR except in the special case where all variables are assumed to have the same order of fractional integration.

We apply our methodology to study technology shocks. The identification of technology shocks using VARs has generated some controversy — and the central point of contention is the modeling of low-frequency patterns. Building on Blanchard and Quah (1989), Galí (1999) uses long-run restrictions in a VAR: He argues that total-factor productivity (TFP) shocks are the only shocks that have permanent effects on output per hour. In stark contrast to the predictions of real-business-cycle theory, Galí finds that a positive technology shock has a negative effect on hours of work. Christiano et al. (2003) approach the same problem with the same identifying assumptions, but reach the opposite conclusion. Apparently, if hours enter the VAR in differences (Galí's preferred specification), then positive TFP shocks cause hours to go down; if hours enter the VAR in levels (Christiano et al.'s preferred specification), then positive TFP shocks cause hours to go up.

In other words, a VAR's implications for technology shocks hinge critically on the order of integration.[7] A FIVAR is therefore a natural alternative: We can estimate the VAR parameters simultaneously with each variable's order of integration, without constraining it to be exactly zero or one. Indeed, our empirical results suggest that several of our models' variables are fractionally integrated. Gil-Alana and Moreno (2009) and Lovcha and Perez-Laborda (2015) also investigate technology shocks with FIVARs, but our work differs from theirs in several key respects. First, their papers approach an interesting macroeconomic question using well-established frequentist methods, whereas we introduce a new Bayesian method. Frequentist estimators can produce imprecise estimates with large systems, and those authors focus on bivariate models, although Lovcha and Perez-Laborda also fit a three-variable model as robustness check. Second, because our method is better suited to large systems, we analyze more than one type of technology shock. Following Fisher (2006), we identify investment-specific technology (IST) shocks, as well as TFP shocks. It's assumed that only IST shocks have a permanent effect on the relative price of investment goods, but both IST and TFP and shocks have permanent effects on output per hour. The FIVAR suggests that an increase in TFP leads to an initial drop in hours, regardless of whether hours enter in levels or differences. For IST shocks, the FIVAR and the VAR produce substantially different impulse responses. Following a decline in IST, the VAR predicts an immediate drop in output driven entirely by a drop in hours; the FIVAR predicts a delayed decline in output coming from a gradual fall in productivity, with hours staying nearly constant.

We will proceed as follows. Section 2 introduces the FIVAR model. We outline our estimation strategy in Section 3, and we discuss sampling algorithms in Section 4. Section 5 contains the empirical application on technology shocks. Section 6 concludes. All proofs are in Appendix A; supplementary results are in

[7] More generally, the results depend on the specification of low-frequency movements. See Fernald (2007) for a discussion. Francis and Ramey (2009) advocate using an alternative measure of hours that accounts for the demographic composition of the labor force. Effectively, accounting for slow-moving demographic trends alters the low-frequency properties of the data, and those authors find that positive TFP shocks depress hours.

Appendix B. Throughout, we will use an asterisks (*) to denote the conjugate transpose, and a prime ($'$) to denote the standard (non-conjugate) transpose.

2 The FIVAR Model

For an $n \times 1$ vector $\mathbf{x}_t \equiv (x_{1,t}, \ldots, x_{n,t})'$, a Gaussian FIVAR process assumes the form:

$$\mathbf{A}(L)\,\mathbf{D}(L)\,\mathbf{x}_t = \mathbf{e}_t, \quad \mathbf{e}_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \mathbf{Q}_e^{-1}\right), \tag{2.1}$$

where L is the lag operator, $\mathbf{A}(L)$ is an m^{th}-order lag polynomial, and $\mathbf{D}(L)$ is a diagonal matrix of fractional-differencing operators:

$$\mathbf{A}(L) \equiv \mathbf{I}_n - \sum_{\ell=1}^{m} \mathbf{A}_\ell L^\ell, \quad \mathbf{D}(L) \equiv \mathrm{diag}\left((1-L)^{\delta_1}, \cdots, (1-L)^{\delta_n}\right). \tag{2.2}$$

That is, $\mathbf{D}(L)\,\mathbf{x}_t$ follows a VAR(m), and the j^{th} element of $\mathbf{D}(L)\,\mathbf{x}_t$ is $(1-L)^{\delta_j} x_{j,t}$. If δ_j is an integer, then $(1-L)^{\delta_j} x_{j,t}$ is just the δ_j^{th} difference of $x_{j,t}$, and we can write the differencing operator in terms of its binomial expansion:

$$(1-L)^{\delta_j} = \sum_{\ell=0}^{\delta_j} (-1)^\ell \binom{\delta_j}{\ell} L^\ell. \tag{2.3}$$

If δ_j is not an integer, then this operator is defined directly in terms of its binomial expansion, replacing the factorial with the Gamma function:

$$(1-L)^{\delta_j} = \sum_{\ell=0}^{\infty} \frac{\Gamma(\ell - \delta_j)}{\Gamma(\ell+1)\,\Gamma(-\delta_j)} L^\ell. \tag{2.4}$$

Now, with δ_j no longer an integer, this polynomial contains an infinite number of lags. For $\mathbf{x}_t$ to be stationary and invertible, it's necessary that each δ_j falls in the interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$.[8] The series $x_{j,t}$ is said to exhibit long memory if $\delta_j \in \left(0, \frac{1}{2}\right)$. In that case, the autocovariance function decays harmonically — i.e., $\mathbb{E}\left[x_{j,t} x_{j,t-\ell}\right] \sim O\left(|\ell|^{2\delta_j - 1}\right)$ — much more slowly than the exponential decay that arises in a standard VAR model. Long-memory processes are dominated by slow-moving oscillations: At frequency zero, the power spectral density of $x_{j,t}$ is infinite. If $\delta_j \in \left(-\frac{1}{2}, 0\right)$, then $x_{j,t}$ is said to exhibit negative memory. Then, the autocovariance function decays quickly, and the power spectral density is zero at frequency zero. Negative memory can arise if, for instance, a series has been over-differenced.

Evidently, a FIVAR can accommodate low-frequency dynamics that depart significantly from a VAR. The

[8] As with a VAR, stationarity for the FIVAR also requires all roots of the polynomial $\det(\mathbf{A}(z))$ to fall outside the unit circle.

cost of this flexibility, in terms of parsimony, is relatively small. When fitting a Bayesian VAR, Litterman (1986) suggests including "as long a lag as is computationally feasible, with a prior distribution on the coefficients reflecting the fact that coefficients on longer lags are more likely to be close to zero." (p. 28) In practice, this principle often translates into a year's worth of lags, plus one. With five series observed on a quarterly basis, a VAR would contain 135 parameters; a FIVAR would require just five more. The differencing parameters control how quickly the coefficients in the lag polynomial (2.4) decay to zero. In that sense, adopting an informative prior over the differencing parameters is one way of implementing Litterman's advice: We're allowing $\mathbf{x}_t$ to depend on its own infinite history, but controlling the degree of dependence.

3 Estimation

First, in Section 3.1, we will discuss Whittle's approximation to the likelihood, and the associated quasi-posterior, for a general stationary time-series model. Then, in Section 3.2, we will apply this approximation to the FIVAR model specifically. These analytical results set the stage for the sampling algorithms presented in Section 4. Throughout, we will use the notation $p(\cdot)$ to refer to a generic probability density.

3.1 A Frequency-Domain Posterior Approximation

Let $\mathbf{x}_t$ be an $n \times 1$ vector of mean-zero stationary time series. Suppose that $\mathbf{x}^T \equiv \{\mathbf{x}_t\}_{t=0}^{T-1}$ are jointly Gaussian, and the autocovariance function of $\mathbf{x}_t$ is parameterized by a collection of parameters θ. Let $L\left(\mathbf{x}^T \mid \theta\right)$ be the associated likelihood function. Given a prior density $p(\theta)$, the posterior density for θ is given by:

$$p\left(\theta \mid \mathbf{x}^T\right) = \frac{L\left(\mathbf{x}^T \mid \theta\right) p(\theta)}{\int L\left(\mathbf{x}^T \mid \acute{\theta}\right) p\left(\acute{\theta}\right) d\acute{\theta}}. \tag{3.1}$$

A common strategy is to replace the exact time-domain likelihood function with Whittle's frequency-domain approximation. Denote the discrete Fourier transform (DFT) of $\mathbf{x}^T$ by $\mathbf{z}^T \equiv \{\mathbf{z}_k\}_{k=0}^{T-1}$:

$$\mathbf{z}_k \equiv \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \mathbf{x}_t \exp\left\{-i\omega_k t\right\}, \tag{3.2}$$

where $\omega_k \equiv \frac{2\pi k}{T}$ is the k^{th} Fourier frequency. Denote by $K \equiv \lfloor T/2 - 1 \rfloor$ be the number of non-redundant Fourier frequencies.[9] Whittle's approximation takes the following form:

$$L\left(\mathbf{x}^T \mid \theta\right) \approx \hat{L}\left(\mathbf{z}^T \mid \theta\right) \equiv \prod_{k=1}^{K} \pi^{-n} \det\left(\mathbf{f}\left(\omega_k \mid \theta\right)\right)^{-1} \exp\left\{-\mathbf{z}_k^* \mathbf{f}\left(\omega_k \mid \theta\right)^{-1} \mathbf{z}_k\right\}, \tag{3.3}$$

where $\mathbf{f}\left(\omega \mid \theta\right)$ is the spectral density, which is also parameterized by θ. We will refer to $L\left(\mathbf{x}^T \mid \theta\right)$ as the time-domain likelihood, and $\hat{L}\left(\mathbf{z}^T \mid \theta\right)$ as the frequency-domain likelihood. Likewise, we will refer to $p\left(\theta \mid \mathbf{x}^T\right)$ as the time-domain posterior, and we will define the frequency-domain posterior as:

$$\hat{p}\left(\theta \mid \mathbf{x}^T\right) \equiv \frac{\hat{L}\left(\mathbf{x}^T \mid \theta\right) p\left(\theta\right)}{\int \hat{L}\left(\mathbf{x}^T \mid \acute{\theta}\right) p\left(\acute{\theta}\right) d\acute{\theta}}. \tag{3.4}$$

Strictly speaking, only $p\left(\theta \mid \mathbf{x}^T\right)$ represents the exact posterior distribution, and $\hat{p}\left(\theta \mid \mathbf{x}^T\right)$ is a quasi-posterior approximation. However, this formulation is common in the Bayesian time-series literature, because in many applications, evaluating the frequency-domain likelihood requires much less computing time (and less coding time) than evaluating the time-domain likelihood.[10] Certainly, that's true for the FIVAR model: The exact likelihood requires calculating the full $Tn \times Tn$ variance-covariance matrix, plus its inverse and determinant. Nevertheless, to make the sampling problem tractable, we need to do more than simply replace $L\left(\mathbf{x}^T \mid \theta\right)$ with $\hat{L}\left(\mathbf{x}^T \mid \theta\right)$. In a FIVAR model, θ is typically high-dimensional, so even if the frequency-domain posterior is less costly to evaluate, it may still be difficult to maximize or integrate. We will therefore turn our attention to the specific features of the FIVAR model, and how they relate to the computationally simpler VAR model.

3.2 Approximating the FIVAR Posterior

Assume that $\mathbf{x}_t$ follows a FIVAR (2.1). Our goal is to form beliefs about the parameters $\theta \equiv \left(\boldsymbol{\delta}', \text{vech}\left(\mathbf{Q}_e\right)', \mathbf{a}'\right)'$, where we have defined:

$$\boldsymbol{\delta} \equiv \left(\delta_1, \ldots, \delta_n\right)', \quad \mathbf{a} \equiv \text{vec}\left(\mathbf{A}\right), \quad \mathbf{A} \equiv \begin{bmatrix} \mathbf{A}_1 & \cdots & \mathbf{A}_m \end{bmatrix}. \tag{3.5}$$

Let $\mathbf{f}_{FIVAR}\left(\omega \mid \theta\right)$ denote the spectral density of a FIVAR, and let $\mathbf{f}_{VAR}\left(\omega \mid \theta\right)$ denote the spectral density of a VAR. By construction, a FIVAR process comes from applying a fractional-integration filter to a VAR

[9] Half of the DFT ordinates are redundant, in the sense that $\mathbf{z}_k' = \mathbf{z}_{T-k}^*$.

[10] For example, in the econometrics literature, Plagborg-Møller (2016) and Sala (2015) use the frequency-domain posterior (3.4) to fit structural time-series models without long memory. In the statistics literature, Liseo et al. (2001) and Holan et al. (2009) use the frequency-domain posterior for Bayesian estimation of univariate long-memory models. Tamaki (2008) proves a version of the Bernstein-von Mises theorem for the frequency-domain posterior with univariate long-memory processes.

process. Hence:

$$\mathbf{f}_{FIVAR}(\omega \mid \theta) = \mathbf{D}(\exp\{-i\omega\})^{-1} \mathbf{f}_{VAR}(\omega \mid \theta) \mathbf{D}(\exp\{-i\omega\})^{-1*}. \tag{3.6}$$

This fact implies a relationship between the frequency-domain likelihood for a FIVAR process, denoted $\hat{L}_{FIVAR}(\mathbf{z}^T \mid \theta)$, and the frequency-domain likelihood for a VAR process, denoted $\hat{L}_{VAR}(\mathbf{z}^T \mid \theta)$.

Proposition 1. *Let* $\tilde{\mathbf{z}}^T \equiv \{\tilde{\mathbf{z}}_k\}_{k=0}^{T-1}$ *with* $\tilde{\mathbf{z}}_0 \equiv \mathbf{0}_{n\times 1}$ *and* $\tilde{\mathbf{z}}_k \equiv \mathbf{D}(\exp\{-i\omega_k\})\mathbf{z}_k$, $k \geq 1$. *We can write the frequency-domain likelihood for the FIVAR process as:*

$$\hat{L}_{FIVAR}(\mathbf{z}^T \mid \theta) = \kappa^{\mathbf{1}'\boldsymbol{\delta}} \hat{L}_{VAR}(\tilde{\mathbf{z}}^T \mid \theta), \tag{3.7}$$

where $\kappa \equiv \prod_{k=1}^{K} [2 - 2\cos(\omega_k)]$.

Proposition 1 suggests a strategy for marginalizing over the VAR parameters to approximate the posterior distribution of the differencing parameters. For a given $\boldsymbol{\delta}$, let $\tilde{\mathbf{x}}^T \equiv \{\tilde{\mathbf{x}}_t\}_{t=0}^{T-1}$ be defined as the inverse DFT of $\tilde{\mathbf{z}}^T$:

$$\tilde{\mathbf{x}}_t \equiv \frac{1}{\sqrt{T}} \sum_{k=0}^{T-1} \tilde{\mathbf{z}}_k \exp\{i\omega_k t\}. \tag{3.8}$$

Applying equation (3.3) again, we will approximate the frequency-domain VAR likelihood with the corresponding time-domain VAR likelihood:[11]

$$\begin{aligned} \hat{L}_{VAR}(\tilde{\mathbf{z}}^T \mid \theta) &\approx L_{VAR}(\tilde{\mathbf{x}}^T \mid \theta) \\ &= \prod_{t=m}^{T-1} (2\pi)^{-\frac{n}{2}} \det(\mathbf{Q}_e)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\tilde{\mathbf{x}}_t - \sum_{\ell=1}^{m} \mathbf{A}_\ell \tilde{\mathbf{x}}_{t-\ell}\right)' \mathbf{Q}_e \left(\tilde{\mathbf{x}}_t - \sum_{\ell=1}^{m} \mathbf{A}_\ell \tilde{\mathbf{x}}_{t-\ell}\right)\right\}. \end{aligned} \tag{3.9}$$

Combining equations (3.7) and (3.9) gives us:

$$\hat{L}_{FIVAR}(\mathbf{z}^T \mid \theta) \approx \kappa^{\mathbf{1}'\boldsymbol{\delta}} L_{VAR}(\tilde{\mathbf{x}}^T \mid \theta) \equiv \tilde{L}_{FIVAR}(\tilde{\mathbf{x}}^T \mid \theta). \tag{3.10}$$

We will refer to $\tilde{L}_{FIVAR}(\tilde{\mathbf{x}}^T \mid \theta)$ as the hybrid likelihood, because it combines the frequency-domain FIVAR likelihood with the time-domain VAR likelihood. In turn, we will define the hybrid posterior as:

$$\tilde{p}(\theta \mid \mathbf{x}^T) \equiv \frac{\tilde{L}_{FIVAR}(\tilde{\mathbf{x}}^T \mid \theta) p(\theta)}{\int \tilde{L}_{FIVAR}(\tilde{\mathbf{x}}^T \mid \acute{\theta}) p(\acute{\theta}) d\acute{\theta}}. \tag{3.11}$$

Like the frequency-domain posterior, the hybrid posterior is only an approximation to the exact, time-domain posterior. Nevertheless, this formulation has a clear interpretation, and it will be much easier to sample from

[11] This formulation of the time-domain VAR likelihood conditions on the first m observations.

$\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ than from $p\left(\theta \mid \mathbf{x}^T\right)$. Recall that, to construct $\tilde{\mathbf{z}}_k$, we multiplied the Fourier transform of the data $\mathbf{z}_k$ by the transfer function associated with a fractional-differencing filter $\mathbf{D}\left(\exp\left\{-i\omega_k\right\}\right)$. Consequently, $\tilde{\mathbf{x}}_t$ is the finite-sample analogue of $\mathbf{D}\left(L\right)\mathbf{x}_t$, which is a VAR process.[12] Likewise, for a fixed value of $\boldsymbol{\delta}$, the hybrid posterior $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ is proportional to $L_{VAR}\left(\tilde{\mathbf{x}}^T \mid \theta\right) p\left(\theta\right)$. That is, conditional on $\boldsymbol{\delta}$, forming beliefs about $\left(\mathbf{Q}_e, \mathbf{a}\right)$ using the hybrid posterior is equivalent to fitting a Bayesian VAR, taking $\tilde{\mathbf{x}}^T$ as data. This fact is computationally convenient, because it implies a semi-conjugate prior, as demonstrated by the following proposition.

Proposition 2. *Assume that the prior for $\left(\mathbf{Q}_e, \mathbf{a}\right)$ is independent of $\boldsymbol{\delta}$ and takes a normal-Wishart form:*

$$\begin{aligned} \mathbf{Q}_e &\sim \mathrm{W}\left(\frac{1}{\nu}\bar{\mathbf{Q}}_e, \nu\right) && (3.12) \\ \mathbf{a} \mid \mathbf{Q}_e &\sim \mathrm{N}\left(\mathrm{vec}\left(\bar{\mathbf{A}}\right), \mathbf{Q}_a^{-1} \otimes \mathbf{Q}_e^{-1}\right). && (3.13) \end{aligned}$$

For a given value of $\boldsymbol{\delta}$, let $\tilde{\mathbf{x}}_t$ be defined as in equation (3.8). Define:

$$\tilde{\mathbf{Y}} \equiv \left[\begin{array}{cccc} \tilde{\mathbf{x}}_m & \tilde{\mathbf{x}}_{m+1} & \cdots & \tilde{\mathbf{x}}_{T-1} \end{array}\right]', \quad \tilde{\mathbf{X}}_t \equiv \left[\begin{array}{cccc} \tilde{\mathbf{x}}_{t-1}' & \tilde{\mathbf{x}}_{t-2}' & \cdots & \tilde{\mathbf{x}}_{t-m}' \end{array}\right]', \quad \tilde{\mathbf{X}} \equiv \left[\begin{array}{ccc} \tilde{\mathbf{X}}_p & \cdots & \tilde{\mathbf{X}}_{T-1} \end{array}\right]'.$$

Also, for a given $\boldsymbol{\delta}$, define:

$$\begin{aligned} \hat{\mathbf{Q}}_a &\equiv \mathbf{Q}_a + \tilde{\mathbf{X}}'\tilde{\mathbf{X}} && (3.14) \\ \hat{\mathbf{Q}}_e &\equiv \left(\nu + T - m\right)\left(\nu\bar{\mathbf{Q}}_e^{-1} + \bar{\mathbf{A}}\mathbf{Q}_a\bar{\mathbf{A}}' + \tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} - \hat{\mathbf{A}}\hat{\mathbf{Q}}_a\hat{\mathbf{A}}'\right)^{-1} && (3.15) \\ \hat{\mathbf{A}} &\equiv \left(\bar{\mathbf{A}}\mathbf{Q}_a + \tilde{\mathbf{Y}}'\tilde{\mathbf{X}}\right)\hat{\mathbf{Q}}_a^{-1}. && (3.16) \end{aligned}$$

Under the hybrid posterior, the marginal density of $\boldsymbol{\delta}$ is proportional to:

$$\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right) \propto p\left(\boldsymbol{\delta}\right) \kappa^{\mathbf{1}'\boldsymbol{\delta}} \det\left(\hat{\mathbf{Q}}_a\right)^{-\frac{n}{2}} \det\left(\hat{\mathbf{Q}}_e\right)^{\frac{\nu+T-m}{2}}, \qquad (3.17)$$

and the conditional distribution of $\left(\mathbf{Q}_e, \mathbf{a}\right)$, given $\boldsymbol{\delta}$, is:

$$\begin{aligned} \mathbf{Q}_e \mid \boldsymbol{\delta}, \mathbf{x}^T &\sim \mathrm{W}\left(\frac{1}{\nu + T - m}\hat{\mathbf{Q}}_e, \nu + T - m\right) && (3.18) \\ \mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta}, \mathbf{x}^T &\sim \mathrm{N}\left(\mathrm{vec}\left(\hat{\mathbf{A}}\right), \hat{\mathbf{Q}}_a^{-1} \otimes \mathbf{Q}_e^{-1}\right). && (3.19) \end{aligned}$$

[12] The operator $\mathbf{D}\left(L\right)$ is an infinite-order lag polynomial, so we can't calculate $\mathbf{D}\left(L\right)\mathbf{x}_t$ exactly using a finite sample.

Additionally, the conditional distribution of $\mathbf{Q}_e$*, given* $(\boldsymbol{\delta}, \mathbf{a})$*, is:*

$$\mathbf{Q}_e \mid \mathbf{a}, \boldsymbol{\delta}, \mathbf{x}^T \sim \mathrm{W}\left(\frac{1}{\nu + T + (n-1)\,m}\tilde{\mathbf{Q}}_e, \nu + T + (n-1)\,m\right) \tag{3.20}$$

$$\tilde{\mathbf{Q}}_e \equiv [\nu + T + (n-1)\,m]\left[\left(\frac{1}{\nu + T - m}\hat{\mathbf{Q}}_e\right)^{-1} + \left(\mathbf{A} - \hat{\mathbf{A}}\right)\hat{\mathbf{Q}}_a\left(\mathbf{A} - \hat{\mathbf{A}}\right)'\right]^{-1}. \tag{3.21}$$

Proposition 2 demonstrates that the hybrid posterior factors pleasantly, which will allow us to break up the problem into manageable components. The only non-standard part of the hybrid posterior is the marginal distribution of $\boldsymbol{\delta}$. Fortunately, the dimension of $\boldsymbol{\delta}$ is only n. Conversely, the VAR parameters are high-dimensional, but their conditional distributions are standard. These facts will form the basis of the sampling algorithms in Section 4. The prior independence between $(\mathbf{Q}_e, \mathbf{a})$ and $\boldsymbol{\delta}$ can easily be relaxed, as long the conditional prior for the VAR parameters retains a normal-Wishart form; i.e., one could specify ν, $\bar{\mathbf{Q}}_e$, $\bar{\mathbf{A}}$, and $\mathbf{Q}_a$ as functions of $\boldsymbol{\delta}$.

The structure of the hybrid posterior also makes it relatively easy to find the mode. The mode provides a convenient point estimate during preliminary data exploration, before performing Monte Carlo integration. Also, as we'll see in Section 4.3, knowing the shape of the posterior local to the mode can be helpful for constructing an efficient sampler. Maximizing $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ by brute force would be cumbersome, because it's a high-dimensional optimization problem. Fortunately, we can use the results from Proposition 2 to concentrate out the VAR parameters, so that we only have to maximize over $\boldsymbol{\delta}$. Then, given the maximizing value of $\boldsymbol{\delta}$, it's straightforward to back out the maximizing values of $\mathbf{Q}_e$ and $\mathbf{a}$. Proposition 3 delineates the solution.

Proposition 3. *Let* $\left(\boldsymbol{\delta}^\dagger, \mathbf{Q}_e^\dagger, \mathbf{a}^\dagger\right)$ *be the parameters that maximize the hybrid posterior. Then:*

$$\boldsymbol{\delta}^\dagger = \underset{\boldsymbol{\delta} \in \left[-\frac{1}{2}, \frac{1}{2}\right]^n}{\operatorname{argmax}}\; p(\boldsymbol{\delta})\,\kappa^{\mathbf{1}'\boldsymbol{\delta}} \det\left(\hat{\mathbf{Q}}_e\right)^{\frac{\nu+T+(n-1)m-n-1}{2}}. \tag{3.22}$$

where $\hat{\mathbf{Q}}_e$ *is taken to be a function* $\boldsymbol{\delta}$*, as given by equation (3.15). Given* $\boldsymbol{\delta}^\dagger$*, the maximizing values of* $(\mathbf{Q}_e, \mathbf{a})$ *are given by:*

$$\mathbf{Q}_e^\dagger \equiv \frac{\nu + T - m + (m-1)\,n - 1}{\nu + T - m}\hat{\mathbf{Q}}_e, \quad \mathbf{a}^\dagger \equiv \operatorname{vec}\left(\hat{\mathbf{A}}\right), \tag{3.23}$$

where $\hat{\mathbf{Q}}_e$ *and* $\hat{\mathbf{A}}$ *are calculated using equations (3.15) and (3.16), evaluated using* $\boldsymbol{\delta} = \boldsymbol{\delta}^\dagger$*.*

4 Sampling Algorithms

We will discuss several strategies for sampling from the hybrid posterior $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ and from the time-domain posterior $p\left(\theta \mid \mathbf{x}^T\right)$. Section 4.1 provides a fast and simple way of sampling from the hybrid posterior. In Section 4.2, we use the method from Section 4.1 to form an efficient proposal distribution for sequential importance sampling, which provides a tractable way of generating draws from the exact, time-domain posterior. Section 4.3 describes a strategy for Metropolis sampling, which is used as an intermediate step in the sequential importance sampler. Section 4.4 provides an efficient method for computing the marginal data density. Section 4.5 provides some practical guidelines for implementing the algorithms.

4.1 Sampling from the Hybrid Posterior

Proposition 2 suggests a strategy for sampling θ from the hybrid posterior:

1. Draw $\boldsymbol{\delta}$ from $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$.
2. Given $\boldsymbol{\delta}$, draw $\mathbf{Q}_e \sim \mathrm{W}\left(\frac{1}{\nu+T-m}\hat{\mathbf{Q}}_e, \nu+T-m\right)$.
3. Given $\boldsymbol{\delta}$ and $\mathbf{Q}_e$, draw $\mathbf{a} \sim \mathrm{N}\left(\mathrm{vec}\left(\hat{\mathbf{A}}\right), \hat{\mathbf{Q}}_a^{-1} \otimes \mathbf{Q}_e^{-1}\right)$.

Steps 2 and 3 are trivial, so we just need a way to draw from the non-standard distribution $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$. We've found that Metropolis-Hastings is fast and works well. We recommend using a proposal distribution that mixes between an independent proposal from a uniform distribution and a random-walk proposal form a normal distribution. That is, if $\boldsymbol{\delta}^{(i)}$ is the i^{th} draw from the Metropolis chain, then the subsequent draw is proposed from the distribution:

$$\boldsymbol{\delta}^{prop} \mid \boldsymbol{\delta}^{(i)} \sim \begin{cases} \mathrm{Uniform}\left(-\frac{1}{2}, \frac{1}{2}\right)^n & \text{with probability } \upsilon \\ \mathrm{N}\left(\boldsymbol{\delta}^{(i)}, \boldsymbol{\Sigma}_\delta\right) & \text{with probability } 1-\upsilon. \end{cases} \tag{4.1}$$

The corresponding proposal density, denoted $q\left(\cdot \mid \cdot\right)$, is:

$$q\left(\boldsymbol{\delta}^{prop} \mid \boldsymbol{\delta}^{(i)}\right) = \upsilon \mathbb{I}\left[\boldsymbol{\delta}^{prop} \in \left(-\frac{1}{2}, \frac{1}{2}\right)^n\right] + (1-\upsilon)\,\varphi\left(\boldsymbol{\delta}^{prop} \mid \boldsymbol{\delta}^{(i)}, \boldsymbol{\Sigma}_\delta\right), \tag{4.2}$$

where $\varphi\left(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ is the density of the multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. The variance of the random-walk component is motivated by a Laplace approximation to the hybrid posterior. That is:

$$\boldsymbol{\Sigma}_\delta \equiv \left[\left.\frac{\log\left(\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)\right)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^\circ}\right]^{-1}, \quad \boldsymbol{\delta}^\circ \equiv \underset{\boldsymbol{\delta}\in\left[-\frac{1}{2},\frac{1}{2}\right]^n}{\mathrm{argmax}}\ \tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right). \tag{4.3}$$

Algorithm 1 Sampling from the Hybrid Posterior

1. *Initialization.* Solve the maximization problem in equation (4.3) and compute the associated Hessian. Initialize $\boldsymbol{\delta}^{(0)} = \boldsymbol{\delta}^{\circ}$.

2. *Metropolis sampling of* $\boldsymbol{\delta}$. For $i = 1, 2, \ldots, N_0 + N$:

   (a) With probability υ, draw $\boldsymbol{\delta}^{prop} \sim \text{Uniform}\left(-\frac{1}{2}, \frac{1}{2}\right)^n$. With residual probability, draw $\boldsymbol{\delta}^{prop} \sim \text{N}\left(\boldsymbol{\delta}^{(i-1)}, \boldsymbol{\Sigma}_{\delta}\right)$.

   (b) With probability $\min\left\{1, \frac{\tilde{p}\left(\boldsymbol{\delta}^{prop}|\mathbf{x}^T\right)}{\tilde{p}\left(\boldsymbol{\delta}^{(i-1)}|\mathbf{x}^T\right)} \frac{q\left(\boldsymbol{\delta}^{(i-1)}|\boldsymbol{\delta}^{prop}\right)}{q\left(\boldsymbol{\delta}^{prop}|\boldsymbol{\delta}^{(i)}\right)}\right\}$, set $\boldsymbol{\delta}^{(i)} = \boldsymbol{\delta}^{prop}$. With residual probability, set $\boldsymbol{\delta}^{(i)} = \boldsymbol{\delta}^{(i-1)}$.

3. *Direct sampling of* $(\mathbf{Q}_e, \mathbf{a})$. Discard the burn-in, $\left\{\boldsymbol{\delta}^{(i)}\right\}_{i=0}^{N_0}$. For $i = N_0 + 1, \ldots, N_0 + N$:

   (a) Draw $\mathbf{Q}_e^{(i)} \sim \text{W}\left(\frac{1}{\nu+T-m}\hat{\mathbf{Q}}_e^{(i)}, \nu + T - m\right)$.

   (b) $\mathbf{a}^{(i)} \sim \text{N}\left(\text{vec}\left(\hat{\mathbf{A}}^{(i)}\right), \hat{\mathbf{Q}}_a^{(i)-1} \otimes \mathbf{Q}_e^{(i)-1}\right)$.

The computational burden of this maximization is very low, because $\boldsymbol{\delta}$ is low-dimensional and bounded to a unit hypercube. Notice that $\boldsymbol{\delta}^{\circ}$, defined above, is not the same as $\boldsymbol{\delta}^{\dagger}$, characterized in Proposition 3; the former maximizes the marginal distribution of $\boldsymbol{\delta}$, whereas the latter maximizes the joint distribution.

Algorithm 1 summarizes the procedure for generating N draws from the hybrid posterior. Notice that these draws from the hybrid posterior are not constrained to make the model stationary. To ensure that $\boldsymbol{\delta}$ is in the stationary region of the parameter space, all we have to do is assume that the prior $p(\boldsymbol{\delta})$ has support $\left[-\frac{1}{2}, \frac{1}{2}\right]^n$. We will also assume that the prior for $\mathbf{a}$ is Gaussian, as in equation (3.13), but truncated to the stationary region. In practice, we execute Algorithm 1 to produce draws from the non-truncated hybrid posterior; then, we simply discard the draws that generate non-stationarity.

4.2 Sequential Importance Sampling from the Time-Domain Posterior

We can use sequential importance sampling to perform inference using the exact, time-domain posterior. Chapter 5 of Herbst and Schorfheide (2015) contains a detailed introduction to sequential importance sampling for empirical macroeconomists, with an emphasis on DSGE models; Herbst and Schorfheide (2014) establish additional technical results. Here, we will provide a brief summary, before discussing how to adapt this method to the FIVAR model.

A basic (non-sequential) importance sampler uses draws from a proposal density $\pi_0(\theta)$ and reweights them to obtain draws from a target density $\pi_1(\theta)$. For a function $h(\theta)$, one can approximate the expectation of $h(\theta)$ under the target distribution by generating N draws $\left\{\theta^{(i)}\right\}_{i=1}^{N}$ from the proposal distribution and

computing:

$$\int h(\theta)\,\pi_1(\theta)\,d\theta \approx \frac{1}{N}\sum_{i=1}^{N} h\left(\theta^{(i)}\right) W^{(i)}, \quad W^{(i)} \equiv \frac{\pi_1\left(\theta^{(i)}\right)/\pi_0\left(\theta^{(i)}\right)}{\frac{1}{N}\sum_{j=1}^{N}\pi_1\left(\theta^{(j)}\right)/\pi_0\left(\theta^{(j)}\right)}. \tag{4.4}$$

Geweke (1989) provides conditions under which the Monte Carlo approximation converges as $N \to \infty$. Given a collection of draws $\left\{\theta^{(i)}\right\}_{i=1}^{N}$ from the proposal distribution, multinomial resampling entails constructing a new collection of draws $\left\{\theta^{(k)}\right\}_{k=1}^{N}$ by setting $\theta^{(k)} = \theta^{(i)}$ with probability $W^{(i)}$. The marginal distribution of these resampled draws is the target distribution.[13] One way to measure the efficiency of the algorithm is with the effective sample size, defined as:

$$ESS \equiv \frac{N}{\frac{1}{N}\sum_{i=1}^{N}\left(W^{(i)}\right)^2}. \tag{4.5}$$

If the proposal distribution were identical to the target distribution, then each $W^{(i)}$ would be equal to one, and the effective sample size would be N, the number of proposals. Conversely, any discrepancy between the proposal and the target densities creates variance in the importance weights, which causes ESS to fall. In general, the efficiency of importance sampling tends to decline with the size of θ, and in high-dimensional settings, even minute differences between the proposal and the target can cause the importance weights to degenerate. For the FIVAR model, one approach would be to use the hybrid posterior as the proposal distribution, and the time-domain posterior as the target. In our experience, importance sampling performs admirably for small systems — say, $n = 2$ or $n = 3$ — because the proposal is designed to emulate the target. Still, basic importance sampling becomes less robust as the number of parameters grows.

We will therefore turn our attention to sequential importance sampling. The idea is to construct a succession of distributions $\{\pi_j(\theta)\}_{j=0}^{N_\tau}$ as a bridge between the initial proposal distribution $\pi_0(\theta)$ and the final target distribution $\pi_{N_\tau}(\theta)$. If $\pi_j(\theta)$ is close to $\pi_{j+1}(\theta)$, then we can perform importance sampling and multinomial resampling, using $\pi_j(\theta)$ as a proposal and $\pi_{j+1}(\theta)$ as a target. As with basic importance sampling, it's advantageous for $\pi_0(\theta)$ to be close to the final target distribution. That way, only a few bridge distributions will be necessary, and each iteration of resampling will be relatively efficient. For $\tau \in [0,1]$, define:

$$\pi(\theta \mid \tau) \equiv \frac{p\left(\theta \mid \mathbf{x}^T\right)^{\tau}\tilde{p}\left(\theta \mid \mathbf{x}^T\right)^{1-\tau}}{\int p\left(\acute{\theta} \mid \mathbf{x}^T\right)^{\tau}\tilde{p}\left(\acute{\theta} \mid \mathbf{x}^T\right)^{1-\tau} d\acute{\theta}}. \tag{4.6}$$

Notice that $\tau = 0$ corresponds to the hybrid posterior, whereas $\tau = 1$ corresponds to the time-domain posterior. We will perform sequential importance sampling, taking the j^{th} bridge distribution to be $\pi_j(\theta) =$

[13] We mention multinomial resampling for expositional simplicity. In practice, it's more efficient to use stratified resampling, which we employ to generate the results in Section 5. See Herbst and Schorfheide (2015) for a discussion.

Algorithm 2 Sequential Importance Sampling from the Time-Domain Posterior

1. *Initialization.* Generate N draws $\left\{\theta_0^{(i)}\right\}_{i=1}^N$ from the hybrid posterior using Algorithm 1. For $i = 1, \ldots, N$, set $W_0^{(i)} = 1$.
2. *Recursion.* For $j = 1, \ldots, N_\tau$:
   (a) *Correction.* Compute the incremental importance weights for each draw i and renormalize:
   $$w_j^{(i)} = W_{j-1}^{(i)} \left[\frac{L_{FIVAR}\left(\mathbf{x}^T \mid \theta_{j-1}^{(i)}\right)}{\tilde{L}_{FIVAR}\left(\mathbf{x}^T \mid \theta_{j-1}^{(i)}\right)}\right]^{\tau_j - \tau_{j-1}}, \quad \bar{W}_j^{(i)} = \frac{w_j^{(i)}}{\frac{1}{N}\sum_k w_j^{(k)}}. \tag{4.7}$$
   (b) *Selection.* Given the importance weights $\left\{\bar{W}_j^{(i)}\right\}_{i=1}^N$, compute ESS_j using equation (4.5).
      i. If $ESS_j < \frac{N}{2}$ or $j = N_\tau$, then perform multinomial resampling — i.e., for each i, set $\acute{\theta}_j^{(i)} = \theta_{j-1}^{(k)}$ with probability $\frac{1}{N}\bar{W}_j^{(k)}$ — and reset $W_j^{(i)} = 1$ for each i.
      ii. If $ESS_j \geq \frac{N}{2}$, then set $\acute{\theta}_j^{(i)} = \theta_{j-1}^{(i)}$ and $W_j^{(i)} = \bar{W}_j^{(i)}$ for each i.
   (c) *Mutation.* For each i, use $\acute{\theta}_j^{(i)}$ to initialize the Metropolis chain described in Section 4.3. Propagate the Metropolis chain N_{MH} times with invariant distribution $\pi\left(\theta \mid \tau_j\right)$, and set $\theta_j^{(i)}$ equal to the final value of the Metropolis chain.

$\pi\left(\theta \mid \tau_j\right)$, where $0 = \tau_0 < \cdots < \tau_{N_\tau} = 1$ is an increasing sequence of tempering parameters. Algorithm 2 presents the details of the sequential importance sampler; we refer interested readers to Herbst and Schorfheide (2014, 2015) for the technical justification underlying each step. We adopt the notational convention of adding a j subscript to parameters whose marginal distribution is $\pi_j\left(\theta\right)$. To prevent the parameter draws from degenerating, it's necessary to undertake a mutation step in between iterations of importance sampling; doing so requires us to propagate a Metropolis chain that has $\pi\left(\theta \mid \tau_j\right)$ as its invariant distribution. Section 4.3 describes how to exploit the hybrid posterior to construct this Metropolis chain in an easy and efficient way. The choice of bridge distributions is the only difference between our procedure and Herbst and Schorfheide's "generic" sequential importance sampler. Those authors begin by drawing from the prior and gradually introduce sample information by specifying $\pi_j\left(\theta\right) \propto p\left(\theta\right) p\left(\mathbf{x}^T \mid \theta\right)^{\tau_j}$, for τ_j increasing from zero to one. Our strategy, by contrast, starts with an approximation to the exact posterior, and gradually phases out the approximation error.

4.3 An Efficient Proposal for Metropolis Chains

To undertake the mutation step in Algorithm 2, we need to generate a Markov process with invariant distribution $\pi\left(\theta \mid \tau\right)$, as defined in equation (4.6). Taking the hybrid posterior as a proposal distribution, we can use the Metropolis-Hastings algorithm. The details of implementing a Metropolis chain with invariant distri-

bution $\pi(\theta \mid \tau)$ are standard, so we only describe the choice of parameter blocks and proposal distributions. We will use the following blocking scheme:

1. Propose $\boldsymbol{\delta}$, conditional on $\mathbf{Q}_e$, $\mathbf{a}$, and $\mathbf{x}^T$.
2. Propose $\mathbf{Q}_e$, conditional on $\boldsymbol{\delta}$, $\mathbf{a}$, and $\mathbf{x}^T$.
3. Propose $\mathbf{a}$, conditional on $\boldsymbol{\delta}$, $\mathbf{Q}_e$, and $\mathbf{x}^T$.

For block one, we propose the differencing parameters using a random-walk step. That is, if $\boldsymbol{\delta}^{(i)}$ is the i^{th} draw from the Metropolis chain, then the subsequent draw is proposed from the distribution:

$$\boldsymbol{\delta}^{prop} \mid \boldsymbol{\delta}^{(i)} \sim \mathrm{N}\left(\boldsymbol{\delta}^{(i)}, c_\delta \boldsymbol{\Sigma}_{\delta|VAR}\right) \tag{4.8}$$

$$\boldsymbol{\Sigma}_{\delta|VAR} \equiv \left[\frac{\partial^2 \log\left(\tilde{p}\left(\theta \mid \mathbf{x}^T\right)\right)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}\bigg|_{\theta=\theta^\dagger}\right]^{-1}, \tag{4.9}$$

where $c_\delta > 0$ is a scaling factor. The choice of $\boldsymbol{\Sigma}_{\delta|VAR}$ is motivated by taking a Laplace approximation to the hybrid posterior. Unlike $\boldsymbol{\Sigma}_\delta$, which was intended to approximate the marginal variance of $\boldsymbol{\delta}$, $\boldsymbol{\Sigma}_{\delta|VAR}$ is intended to approximate the conditional variance of $\boldsymbol{\delta}$, given the VAR parameters. Computing $\boldsymbol{\Sigma}_{\delta|VAR}$ is relatively easy, given the results from Proposition 3. For blocks two and three, we propose $\mathbf{Q}_e$ and $\mathbf{a}$ from their conditional distributions under the hybrid posterior: We propose $\mathbf{Q}_e$ from the Wishart distribution in equation (3.20), and we propose $\mathbf{a}$ from the normal distribution in equation (3.19). Instead of proposing all of the autoregressive coefficients in a single block, it can be helpful to break up $\mathbf{a}$ into several subvectors. Doing so is straightforward, because the elements of $\mathbf{a}$ are jointly Gaussian.

4.4 Computing the Marginal Data Density

The marginal data density (MDD) provides a means of assessing and comparing models within the Bayesian framework; Kass and Raftery (1995) provide a conceptual overview. The MDD is defined as the probability of observing the data, given the model, having marginalized over all the parameters:

$$p\left(\mathbf{x}^T\right) \equiv \int L\left(\mathbf{x}^T \mid \theta\right) p(\theta)\, d\theta. \tag{4.10}$$

Because the FIVAR has a high-dimensional parameter space, the MDD is a high-dimensional integral. Again, the results from Section 3 provide a computationally feasible way of evaluating equation (4.10). To calculate the MDD for a generic posterior distribution, Gelfand and Dey (1994) propose the modified harmonic mean

estimator:

$$p\left(\mathbf{x}^T\right) \approx \left[\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(\theta^{(i)}\right)}{L\left(\mathbf{x}^T \mid \theta^{(i)}\right)p\left(\theta^{(i)}\right)}\right]^{-1}, \tag{4.11}$$

where $\left\{\theta^{(i)}\right\}_{i=1}^{N}$ are draws from the posterior, and where $f(\cdot)$ is a proper probability density whose support is contained in the parameter space. If $f(\theta)$ were identical to $p\left(\theta \mid \mathbf{x}^T\right)$, then equations (3.1) and (4.10) would imply that equation (4.11) holds exactly. Otherwise, for the Monte Carlo approximation to be a good one, a judicious choice of $f(\cdot)$ is necessary to ensure that the summand in equation (4.11) has finite variance; in particular, θ values that have near-zero posterior probability will make this estimate numerically unstable. Geweke (1999) recommends choosing $f(\cdot)$ to approximate the posterior over some compact subset of the parameter space. Truncating the tails of $f(\cdot)$ helps prevent outliers from causing the summand in equation (4.11) to explode.

For the FIVAR model, a good choice of $f(\cdot)$ is especially important because of the number of parameters involved. Fortunately, we already have a good approximation to the exact posterior, in the form of the hybrid posterior. Appendix B.1 explains in detail how we use a truncated version of the hybrid posterior to construct $f(\cdot)$.

4.5 Practical Considerations

Before moving on to the application, we'll touch on some practical considerations when constructing the sampling routines. The performance and efficiency of our algorithms comes from the accuracy of the approximations that undergird the hybrid posterior. Appendix B.2 explores in detail the proximity between the hybrid posterior and the exact, time-domain posterior. For bivariate models, the hybrid and time-domain posteriors are extremely close. For a fixed T, adding more variables to the system somewhat diminishes the quality of the approximation, yet the results from the hybrid posterior and the time-domain posterior are practically very similar, even for the larger models that we estimate. Consequently, our exact-posterior inferences are not overly sensitive to the choice of tuning parameters in the sequential importance sampler. Below, we outline the numerical settings we use to estimate a model with five variables. Readers who are interested in the empirical results can proceed directly to Section 5.

Let's begin with Algorithm 2. The computational burden of sampling from the exact posterior comes mainly from repeat evaluation of the time-domain likelihood function. For each candidate set of parameters, it's necessary to compute the first T autocovariances of the FIVAR process; then, it's necessary to compute the inverse and determinant of the $Tn \times Tn$ variance-covariance matrix of the full sample $\mathbf{x}^T$. To calculate the autocovariances, we use the multivariate splitting algorithm described by Sela and Hurvich (2009), and to deal with the full-sample variance-covariance matrix, we use the recursive least-squares algorithm described

by Sowell (1989a). In Algorithm 2, most of the likelihood evaluations take place in the mutation step. Fortunately, as emphasized by Herbst and Schorfheide (2014, 2015), the mutation step can be executed in parallel: For bridge distribution j, draw i can be propagated independently of draw $i+1$. Consequently, the algorithm's run time is more sensitive to the number of bridge distributions N_τ than the number of draws N. We set $N_\tau = 10$ and $N = 10,000$. The number of bridge distributions is strikingly small — for other models, it's not unusual to use hundreds or thousands.[14] Generally speaking, sequential importance sampling works well when successive bridge distributions are relatively close to one another. In contexts where the initial proposal distribution is far away from the final target distribution, it helps to introduce information gradually with a large number of bridge distributions. In the context of our problem, though, the hybrid posterior is a good approximation to the exact posterior, so it takes only a few steps to get from $\pi_0(\theta)$ to $\pi_{N_\tau}(\theta)$. For the tempering parameters, we use $\tau_j = j/N_\tau$. One way to assess the efficacy of a particular tempering scheme is to look at ESS_j for each bridge distribution in the recursion: If ESS_j starts out too low, then the bridge distributions are probably too far apart. This is not a problem that we encountered, suggesting that our sequence of bridge distributions is reasonable.[15]

The mutation step in Algorithm 2 propagates the Metropolis chain described in Section 4.3, which requires some tuning parameters. The scaling factor c_δ controls the dispersion of the random-walk proposals for the differencing parameters. To generate the results in Section 5, we set c_δ adaptively, as in Herbst and Schorfheide (2014), to target a good acceptance rate.[16] However, we also found that simply setting $c_\delta = 1$ produces similar results. Proposing $\mathbf{a}$ in several blocks increases the acceptance rate for each proposal, but at the cost of having to evaluate the likelihood more times. In practice, we break up $\mathbf{a}$ into three randomly assigned blocks. The acceptance rates for $\mathbf{Q}_e$ and $\mathbf{a}$ are fairly high, because the Metropolis chain uses the hybrid posterior as a proposal distribution. (If the hybrid posterior were exactly equal to the time-domain posterior, then the Metropolis steps for $\mathbf{Q}_e$ and $\mathbf{a}$ would be equivalent to Gibbs steps, and the acceptance rate would be one.) Consequently, the mutation step in Algorithm 2 does a good job of preventing degeneracy in the sequential importance sampler. With this in mind, we set $N_{MH} = 1$.

Finally, Algorithm 1 is relatively straightforward, but we will mention a few details. In step 2 of the

[14] For example, Herbst and Schorfheide (2014) use $N_\tau = 500$ to estimate the DSGE model from Smets and Wouters (2007), and Bognanni and Herbst (2018) use $N_\tau = 2,000$ for a Markov-switching VAR.

[15] For example, Model IV, which we discuss in Section 5, has five variables, five lags, and 138 parameters. Using $N = 10,000$ parameter draws and $N_\tau = 10$ bridge distributions, the effective sample sizes for the bridge distributions were 9,440, 8,631, 7,644, 6,548, 5,431, 4,243, 9,536, 8,573, 7,329, and 6,029.

[16] That is, for the j^{th} iteration of the recursion, we use $c_{\delta,j}$. We initialize $c_{\delta,0} = 1$. Then, we set:

$$c_{\delta,j} = c_{\delta,j-1}\left(.95 + .05\frac{\exp\left\{16\left(AR_{j-1} - \frac{1}{3}\right)\right\}}{1 + \exp\left\{16\left(AR_{j-1} - \frac{1}{3}\right)\right\}}\right),$$

where AR_{j-1} is the fraction of $\boldsymbol{\delta}$ proposals accepted in iteration $j-1$. That way, the acceptance rate for $\boldsymbol{\delta}$ proposals will gravitate toward $\frac{1}{3}$.

algorithm, the choice of υ governs the amount of autocorrelation in the Metropolis chain with invariant distribution $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$. We set $\upsilon = 0.20$, which usually generates an acceptance rate of about one third when $n = 5$. For step 2 of Algorithm 1, we run the Metropolis chain for $1,500,000$ iterations. After obtaining the $\boldsymbol{\delta}^{(i)}$ draws, we obtain the $\left(\mathbf{Q}_e^{(i)}, \mathbf{a}^{(i)}\right)$ draws, as described in step 3 of Algorithm 1. To truncate $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$ to the stationary region of the parameter space, we discard any draw i such that $\left(\boldsymbol{\delta}^{(i)}, \mathbf{Q}_e^{(i)}, \mathbf{a}^{(i)}\right)$ implies non-stationarity. To allow for a burn-in, we retain only the last $1,000,000$ stationary draws; then, to reduce the correlation across draws, we only retain every 100^{th} value. This leaves us with $10,000$ draws whose marginal distribution is the hybrid posterior, truncated to the stationary region of the parameter space. To cut computing time, notice that step 3 of Algorithm 1 can be executed in parallel.

5 Application: Technology Shocks

5.1 Identification

Following the structural VAR literature, we will assume that the reduced-form forecast errors $\mathbf{e}_t$ are linear combinations of some structural economic shocks ϵ_t:

$$\mathbf{e}_t = \Xi\epsilon_t, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \mathbf{I}_n\right). \tag{5.1}$$

Combining the above with equation (2.1) yields the structural moving-average representation of $\mathbf{x}_t$:

$$\mathbf{x}_t = \mathbf{D}\left(L\right)^{-1}\mathbf{A}\left(L\right)^{-1}\Xi\epsilon_t. \tag{5.2}$$

The h^{th} moving-average coefficient gives us the structural impulse response $\frac{\partial \mathbf{x}_{t+h}}{\partial \epsilon_t'}$. The data are informative about $\mathbf{Q}_e$, and the matrix Ξ must satisfy $\Xi\Xi' = \mathbf{Q}_e^{-1}$. There are many matrices that satisfy this condition, so we need to make an economic argument to identify Ξ. More precisely, to identify the first s shocks, we need to recover the first s columns of Ξ.

Long-run identification, in the mode of Blanchard and Quah (1989), places restrictions on the permanent effects of ϵ_t. Suppose that we're interested in some economic variable y_t, which is assumed to be difference-stationary. We can identify $\epsilon_{1,t}$ as the only shock that has a non-zero (and finite) permanent effect on y_t; the permanent effects of all other shocks on y_t are assumed to be zero. Without loss of generality, let $x_{1,t} = \Delta y_t$. The effect of ϵ_t on y_{t+h} is:

$$\frac{\partial y_{t+h}}{\partial \epsilon_t'} = \frac{\partial}{\partial \epsilon_t'}\left[y_{t-1} + \sum_{k=0}^{h}\Delta y_{t+k}\right] = \sum_{k=0}^{h}\frac{\partial x_{1,t+k}}{\partial \epsilon_t'}. \tag{5.3}$$

Letting $h \to \infty$ shows that the permanent effects of the shocks (on y_t) are the sums of the impulse responses (for $x_{1,t}$). Tschernig et al. (2013) and Lovcha and Perez-Laborda (2015) point out some subtleties of long-run restrictions with FIVARs. In light of equation (5.2), $\lim_{h\to\infty} \frac{\partial y_{t+h}}{\partial \epsilon_t'}$ equals the first row of $\mathbf{D}(1)^{-1}\mathbf{A}(1)^{-1}\Xi$. Recall that the $(1,1)$ element of $\mathbf{D}(L)$ is $(1-L)^{\delta_1}$. This fact implies that δ_1 must be constrained to equal zero: If δ_1 were positive, then the first row of $\mathbf{D}(1)^{-1}\mathbf{A}(1)^{-1}\Xi$ would be infinite, and if δ_1 were negative, then the first row of $\mathbf{D}(1)^{-1}\mathbf{A}(1)^{-1}\Xi$ would be zero. With the restriction $\delta_1 = 0$ in place, it will be more convenient to look at the long-run effect of ϵ_t on $\mathbf{D}(L)\mathbf{x}_t$, keeping in mind that the first element of $\mathbf{D}(L)\mathbf{x}_t$ is simply $x_{1,t} = \Delta y_t$. An argument analogous to the one above shows that:

$$\lim_{h\to\infty} \sum_{k=0}^{h} \frac{\partial \mathbf{D}(L)\mathbf{x}_{t+k}}{\partial \epsilon_t'} = \mathbf{A}(1)^{-1}\Xi. \tag{5.4}$$

One solution is to constrain $\mathbf{A}(1)^{-1}\Xi$ to be lower triangular, which ensures that only the first coordinate of $\lim_{h\to\infty} \frac{\partial y_{t+h}}{\partial \epsilon_t'}$ is non-zero. This restriction implies that $\mathbf{A}(1)^{-1}\Xi$ is the lower Cholesky factor of $\mathbf{A}(1)^{-1}\mathbf{Q}_e^{-1}\mathbf{A}(1)^{-1\prime}$. Hence:

$$\Xi = \mathbf{A}(1)\,\text{chol}\left(\mathbf{A}(1)^{-1}\mathbf{Q}_e^{-1}\mathbf{A}(1)^{-1\prime}\right). \tag{5.5}$$

There are other matrices $\tilde{\Xi}$ such that $\tilde{\Xi}\tilde{\Xi}' = \mathbf{Q}_e^{-1}$, and the first row of $\mathbf{A}(1)^{-1}\tilde{\Xi}$ has a positive number in the first coordinate and zeros everywhere else. However, Christiano et al. (2006) show that all such matrices have the same first column; thus, to identify $\epsilon_{1,t}$, it is sufficient to compute the first column of the matrix Ξ as defined in equation (5.5). More generally, we can can use the first s columns of Ξ to identify s shocks using long-run restrictions: If $x_{j,t} = \Delta y_{j,t}$ and $\delta_j = 0$, then $(\epsilon_{1,t}, \ldots, \epsilon_{j,t})$ will have permanent effects on $y_{j,t}$, but $(\epsilon_{j+1,t}, \ldots, \epsilon_{n,t})$ will not.

From the above formulation, we can see why the distinction between FIVARs and VARs is important for identifying economic shocks with long-run restrictions. If $\mathbf{x}_t$ is generated by a FIVAR with long memory, then the misspecification from fitting a VAR is most severe at low frequencies: For ω local to zero, $\mathbf{f}_{VAR}(\omega \mid \theta)$ is flat, but some elements of $\mathbf{f}_{FIVAR}(\omega \mid \theta)$ diverge to infinity. Equation (5.5) shows why this fact is so relevant for long-run identification: $\mathbf{A}(1)^{-1}\mathbf{Q}_e^{-1}\mathbf{A}(1)^{-1\prime}$ is equal to $\mathbf{f}_{VAR}(0 \mid \theta)$. Estimating the wrong model does more than just degrade our ability to fit the data. By neglecting fractional integration, we might misapprehend the mapping between statistical forecasting errors and structural economic shocks.

We will consider two types of long-run restrictions. First, we will estimate models using the same identification scheme as Galí (1999), who argues that only total-factor productivity (TFP) can have a permanent effect on labor productivity. That is, we specify $x_{1,t}$ as the log difference of real output per hour, and we interpret $\epsilon_{1,t}$ as a TFP shock. Second, we will use the identification scheme introduced by Fisher

(2006), who argues that only investment-specific technology (IST) can have a permanent effect on the price of investment goods, relative to consumption goods. That is, we specify $x_{1,t}$ as the log difference of the relative price of investment (RPI) and $x_{2,t}$ as the log difference of real output per hour; in that case, we interpret $\epsilon_{1,t}$ as an IST shock and $\epsilon_{2,t}$ as a TFP shock. Notice that, in Fisher's framework, both IST and TFP shocks can have permanent effects on labor productivity.

5.2 Data

Table 1 summarizes the model specifications that we consider. We estimate small models (I and II) and larger models (III and IV), allowing hours per capita to enter either in log levels (I and III) or in log differences (II and IV). All four models include labor productivity and hours so that we can identify TFP shocks. Models III and IV include RPI so that we can identify IST shocks as well. In addition, the larger models include inflation and interest rates. Including these variables serves two purposes. First, inflation and interest rates are relevant to investment behavior, so conditioning on these variables can account for some of the variation in RPI. Second, we will be able to see how inflation and interest rates react to technology shocks.

The raw data are available from FRED. Our sample runs from 1955 to 2007; for comparability to other authors, we exclude the Great Recession. The data are quarterly, and all models include five lags. We measure labor productivity growth as 100 times the log difference of real output per hour in the non-farm business sector. We measure hours as 100 times the log of hours of work in the non-farm business sector, minus 100 times the log of the non-institutional population over the age of 16. To construct the price of investment, we compute a chain-weighted average of the NIPA series for gross private domestic investment and durable consumption, and we back out the price deflator. To construct the price of consumption, we compute a chain-weighted average of the NIPA series for nondurable consumption and services consumption, and we back out the price deflator. We measure inflation as 100 times the log difference of our consumption price index. We measure RPI as our investment price index, divided by the consumption price index, and this variable enters our models in log differences, multiplied by 100. The interest rate we consider comes from the 3-month Treasury bill. All series were demeaned before estimation. For each model, we back out the response of log output by summing the responses of log labor productivity and log hours.

5.3 Priors

Our priors are centered on a conventional VAR. Pursuant to the discussion in Section 5.1, identifying technology shocks requires us to impose *a priori* that some elements of $\boldsymbol{\delta}$ are zero. For the unconstrained elements

Table 1: Model Specifications

Model	Variables Included						Shocks Identified	
	RPI Growth	Productivity Growth	Hours (log level)	Hours (log difference)	Inflation Rate	Interest Rate	IST	TFP
I		✓	✓					✓
II		✓		✓				✓
III	✓	✓	✓		✓	✓	✓	✓
IV	✓	✓		✓	✓	✓	✓	✓

of $\boldsymbol{\delta}$, we adopt independent priors:

$$\delta_j + \frac{1}{2} \overset{\text{i.i.d.}}{\sim} \text{Beta}\left(1+\alpha, 1+\alpha\right). \tag{5.6}$$

The choice of a beta distribution confines $\boldsymbol{\delta}$ to the stationary and invertible region. Setting $\alpha = 0$ corresponds to a uniform prior, and letting $\alpha \to \infty$ makes the FIVAR collapse to a VAR. In practice, we set $\alpha = 4$, meaning that $-.25 < \delta_j < .25$ with about 90% probability, and $-.10 < \delta_j < .10$ with about 50% probability. Our priors over $(\mathbf{Q}_e, \mathbf{a})$ are independent of $\boldsymbol{\delta}$ and assume the normal-Wishart form of equations (3.12) and (3.13), truncated to the stationary region of the parameter space. For the choice of hyperparameters, we follow common practices in the Bayesian VAR literature. For $\bar{\mathbf{Q}}_e$, we specify:

$$\bar{\mathbf{Q}}_e = \text{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right)^{-1}, \tag{5.7}$$

where $\hat{\sigma}_j^2$ is the residual variance from a univariate autoregression with variable j.[17] Our prior on $\mathbf{Q}_e$ is relatively diffuse, with $\nu = n+1$. We set $\bar{\mathbf{A}}$ to a matrix of zeros. The autoregressive coefficients are shrunk to zero more aggressively for longer lags:

$$\mathbf{Q}_a = \frac{1}{\lambda^2}\text{diag}\left(1, \ldots, m\right)^2 \otimes \text{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right), \tag{5.8}$$

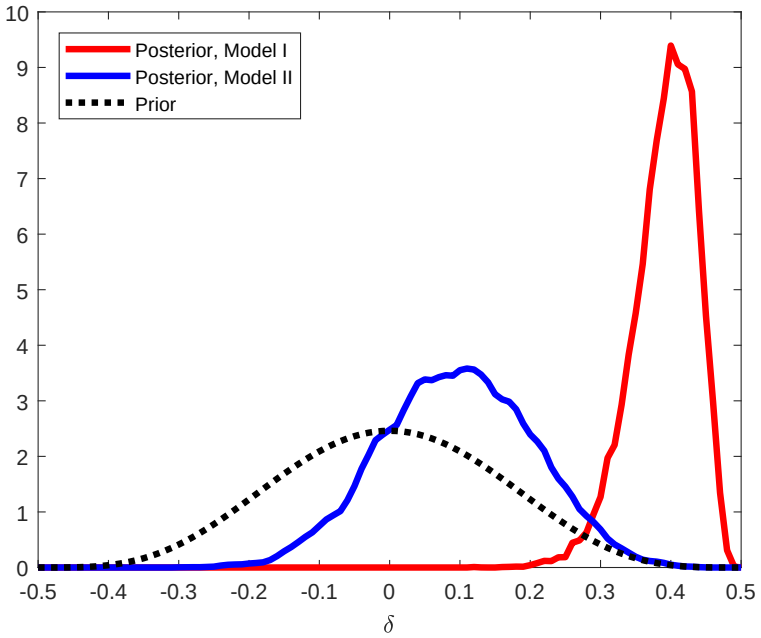
where λ controls the overall tightness of the prior on $\mathbf{a}$. The prior variance of the (i,j) element of $\mathbf{A}_\ell$ is therefore centered at:

$$\mathbb{V}\left[\left[\mathbf{A}_\ell\right]_{i,j} \mid \mathbf{Q}_e = \bar{\mathbf{Q}}_e\right] = \left(\frac{\lambda\hat{\sigma}_i}{\ell\hat{\sigma}_j}\right)^2. \tag{5.9}$$

In practice, we set $\lambda = \frac{1}{5}$.

[17] Strictly speaking, a prior should not contain sample information; however, this is the standard way of eliciting priors in the Bayesian VAR literature.

Figure 1: Differencing Parameters in Models I and II



The posterior distribution of the differencing parameter for the log level of hours (Model I, in red), and the differencing parameter for the log difference of hours (Model II, in blue). The dotted black line is the prior, which is the same across models.

5.4 Bivariate Models

We begin by fitting the smallest models possible to identify TFP shocks. Figure 1 shows the posterior of the differencing parameters for the level of hours (Model I) and the first difference of hours (Model II). The posterior for Model I puts a lot of mass on the level of hours having a differencing parameter near $\frac{1}{2}$, suggesting that a standard VAR with hours in levels would drastically understate the persistence of the series. The posterior from Model II shows weak evidence of long memory in the difference of hours. If hours were over-differenced in Model II, then we would expect to see the posterior in Model II move toward negative memory, rather than long memory. This fact suggests that it's appropriate to take the first difference of hours. More formally, the marginal data density for Model II is about 24 log points higher than the marginal data density for Model I, which is consistent with the graphical diagnostics.

Although the data point toward hours being non-stationary, some macroeconomists might have strong priors that the log level of hours should be stationary. Strictly speaking, hours per capita are bounded, and most theories predict that hours are mean-reverting. However, the assumption that hours are $I(0)$ demands a much stronger prior than the assumption that hours are stationary. As we will see, conditional on treating hours as stationary, estimating Model I with a FIVAR and estimating Model I with a VAR produce substantially different results. But conditional on estimating a FIVAR, Models I and II lead to

Table 2: Variance Decompositions, Bivariate Models

	Model I: FIVAR			Model I: VAR		
	Full Spectrum	Business Cycles	Low Frequency	Full Spectrum	Business Cycles	Low Frequency
Output Growth	11 [3, 32]	7 [1, 27]	23 [11, 45]	60 [21, 92]	57 [19, 91]	70 [33, 96]
Hours Growth	18 [6, 40]	13 [4, 33]	5 [0, 24]	13 [5, 40]	16 [5, 49]	19 [1, 59]
Productivity Growth	66 [39, 89]	64 [36, 86]	94 [82, 99]	91 [71, 96]	84 [65, 93]	83 [61, 96]
	Model II: FIVAR			Model II: VAR		
	Full Spectrum	Business Cycles	Low Frequency	Full Spectrum	Business Cycles	Low Frequency
Output Growth	19 [7, 35]	11 [2, 29]	15 [3, 32]	15 [6, 30]	8 [2, 22]	14 [2, 31]
Hours Growth	9 [4, 21]	8 [3, 18]	2 [0, 12]	12 [5, 25]	10 [4, 22]	3 [0, 14]
Productivity Growth	75 [56, 89]	68 [48, 86]	96 [93, 99]	71 [52, 86]	64 [44, 82]	96 [92, 98]

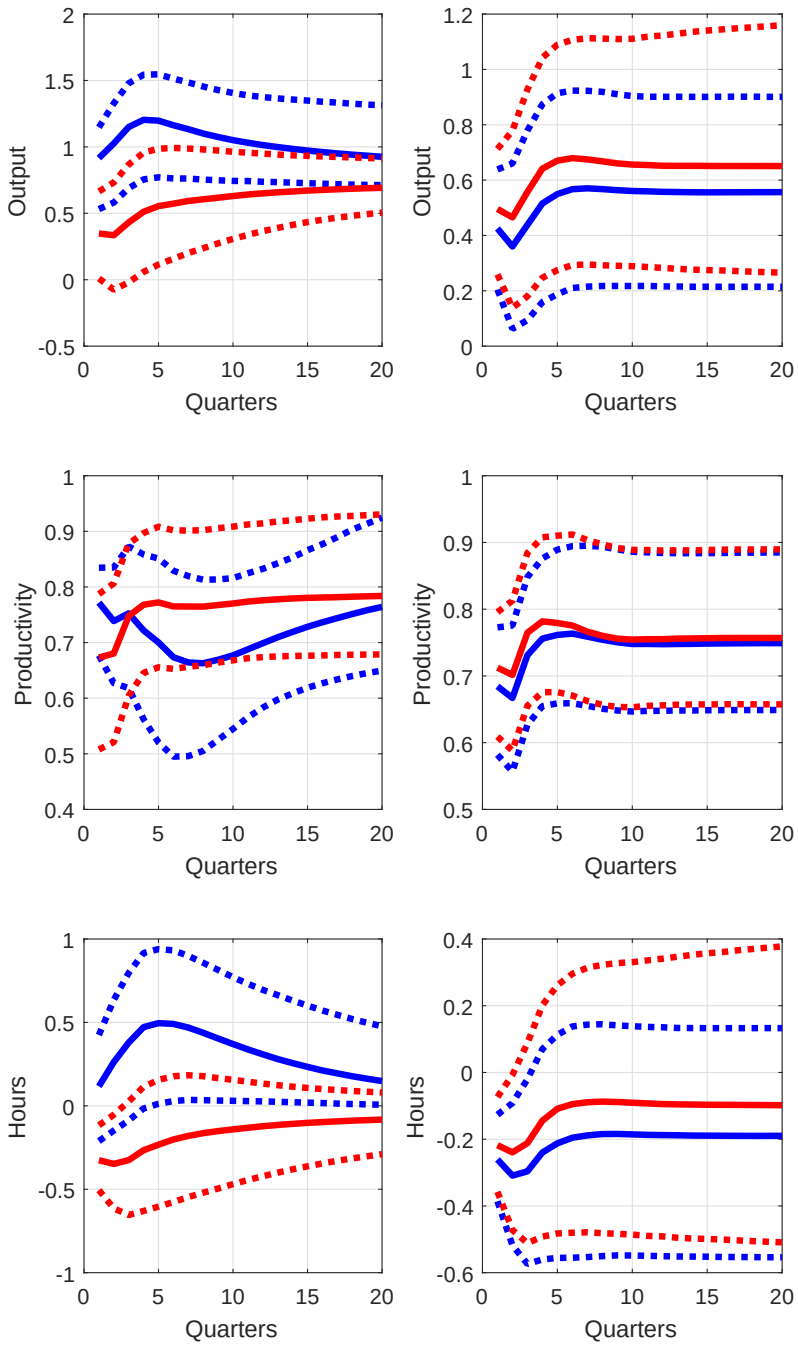
Numbers represent the posterior median of the percent contribution of the identified technology shocks to the variance of each variable coming from oscillations in each frequency band. Numbers in square brackets are the 10th and 90th posterior quantiles. "Full spectrum" refers to the unconditional variance; "business cycles" refers to the variance due to oscillations with periods between 1.5 and 8 years; "low frequency" refers to the variance due to oscillations with periods longer than 8 years. All percentages are rounded to the nearest integer.

similar conclusions about the economic impacts of TFP shocks.

Figure 2 shows the impulse responses from a positive TFP shock in Model I (left column) and in Model II (right column). The blue lines, which represent the results from a standard VAR, tell a story familiar from the literature: When hours enter in levels, a positive TFP shock increases hours of work, but when hours enter in differences, a positive TFP shock decreases hours of work. Like Gil-Alana and Moreno (2009) and Lovcha and Perez-Laborda (2015), we find that the bivariate FIVAR models (red lines) have a more robust prediction: Qualitatively, when TFP goes up, hours go down on impact, regardless of whether hours are assumed to be stationary in levels or stationary in differences. In Model I, the discrepancy between the VAR and the FIVAR is noteworthy, considering that the latter just adds a single parameter to the former. In Model II, the FIVAR impulse responses are quite similar to the VAR impulse responses, reflecting the fact that the posterior for $\boldsymbol{\delta}$ hasn't moved very far from the prior, which is centered at zero.

Table 2 shows the variance decompositions for the bivariate models, estimated using a FIVAR and standard VAR. We have decomposed the variance of the observables by shocks and by frequency bands. We consider the unconditional variance, the variance at business-cycle frequencies (periodicities between 6 and

Figure 2: Impulse Responses from TFP Shocks in Models I and II



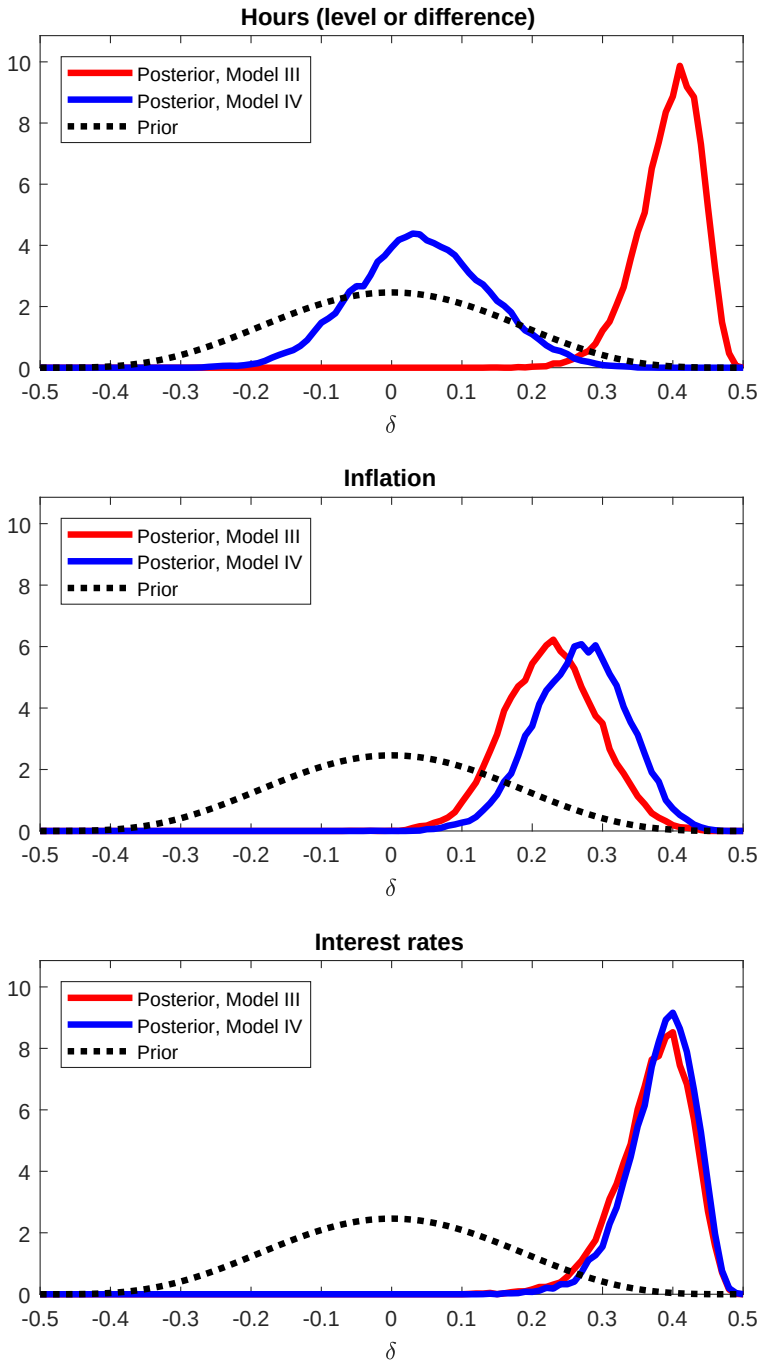
Responses, in levels, from a one-standard-deviation positive shock to TFP. The left column shows results from Model I, with hours in levels; the right column shows results from Model II, with hours in differences. The red lines are results from the FIVAR model; the blue lines are results from a VAR without fractional integration. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

32 quarters), and the variance at low frequencies (periodicities longer than 32 quarters). Let's begin with Model I, summarized in the top panel of Table 2. Whether a practitioner fits Model I with a FIVAR or a VAR determines the measured importance of TFP shocks for driving fluctuations in output growth. The VAR suggests that TFP shocks generate most of the variation in output growth, although the posterior credible sets are quite wide. In contrast, the FIVAR suggests that TFP shocks generate relatively little variation in output growth, especially at business-cycle frequencies. The VAR and the FIVAR both indicate that TFP shocks account for the majority of the variation in productivity growth, but not hours growth. (Even though log hours enter Model I in levels, we have calculated the variance decomposition for the log difference of hours to facilitate a comparison to Model II.) Mechanically, output growth is the sum of hours growth and productivity growth, and the FIVAR impulse responses show that hours and productivity move in opposite directions following a TFP shock. Consequently, according to the FIVAR, TFP generates relatively little movement in output growth. The bottom panel of Table 2 shows the variance decompositions for Model II. Now, the VAR and the FIVAR produce similar results: TFP shocks cause most of the variation in productivity, but not output nor hours. The similarity between the VAR and FIVAR in Model II comes from the fact that the log difference of hours is close to $I(0)$. Interestingly, Models I and II produce similar results when both are estimated with FIVARs, even though they make different assumptions about the stationarity of hours. In a sense, allowing for fractional integration allows the stationary model to be "closer" to the non-stationary model, so the results are less sensitive to ex ante assumptions about stationarity.

5.5 Full Models

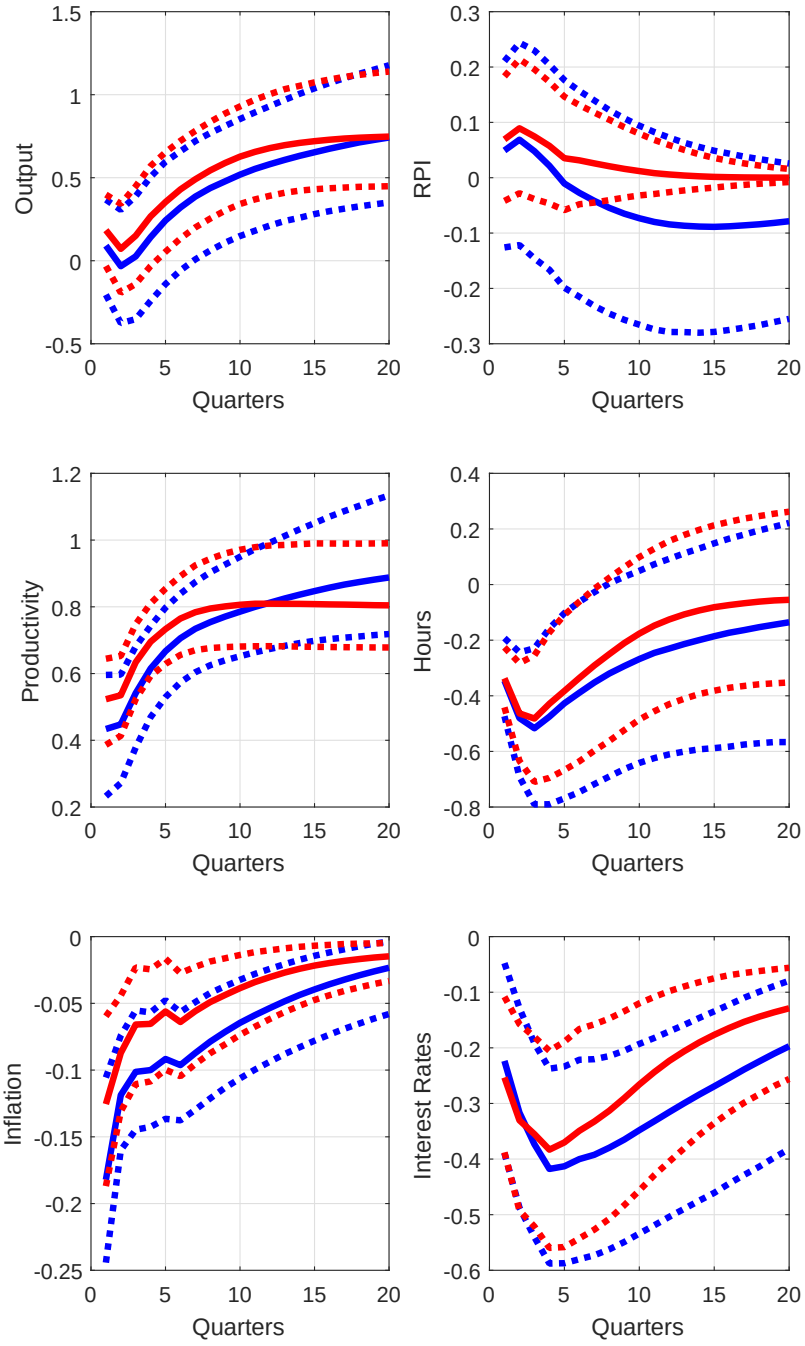
Before studying the identified technology shocks, we will look at the statistical evidence of long memory in the larger models. Models III and IV build on the bivariate models by including RPI, inflation, and interest rates, in addition to labor productivity and hours per capita. Model III renders hours in log levels, and Model IV renders hours in log differences. Recall that, to identify technology shocks, the differencing parameters for the log differences of RPI and labor productivity are constrained to be exactly zero. Figure 3 displays the posterior distribution of unrestricted differencing parameters. Like before, the balance of evidence suggests that it's appropriate to difference hours: The posterior from Model IV suggests that the log difference of hours is nearly $I(0)$, with maybe slight evidence of long memory, and the MDD from Model IV is about 24 log points higher than the MDD from Model III. For inflation and interest rates, both models show strong evidence of long memory, so a standard VAR would struggle to fit the low-frequency fluctuations of these variables. Hereafter, we focus on the results from Model IV, because it's favored by the data; for completeness, the results from Model III are available in Appendix B.3.

Figure 3: Differencing Parameters in Models III and IV



The posterior distribution of the differencing parameters when hours enter in log levels (Model III, in red) and when hours enter in log differences (Model IV, in blue).

Figure 4: Impulse Responses from a TFP Shock in Model IV



Responses from a one-standard-deviation positive shock to TFP in Model VI. The red lines are results from the FIVAR model; the blue lines are results from a VAR without fractional integration. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

TFP Shocks Figure 4 shows the impulse responses from a TFP shock, with the FIVAR depicted in red and the VAR depicted in blue. Given the evidence of long memory, it's interesting how close the FIVAR impulse responses are to the VAR impulse responses. Now, there are two empirical results that run counter to the theory of real business cycles. First, we see the well-known Galí (1999) phenomenon: Hours drop on impact following a positive TFP shock. Second, the theory predicts that higher TFP raises the marginal product of capital, but the impulse responses show that interest rates decline. Note that the drop in nominal interest rates exceeds the drop in inflation, so both real and nominal interest rates fall. It's not obvious what drives this result, but one possibility is the systematic component of monetary policy. Suppose that the central bank can observe output, hours, and inflation — but not the underlying TFP shock. Figure 4 shows that, when the shock hits, hours and inflation both decline. At short horizons, the output response appears to be only slightly positive, and the credible set for the impulse response contains zero. If the central bank were following a Taylor-type rule, then one might expect policymakers to cut interest rates in response to a shock that depresses hours and, to a lesser extent, inflation.

The variance decompositions, shown in Table 3, allow us to assess the relative importance of TFP shocks. Relative to Model II, Model IV ascribes more importance to TFP shocks for explaining hours, but less for output. Adding the additional covariates has a larger affect on the variance decomposition for labor productivity: Although TFP remains important at low frequencies, TFP shocks account for a minority of the variance of labor productivity, either unconditionally or at business-cycle frequencies. Also, TFP shocks contribute very little to observed movements in RPI growth. These conclusions are shared by the FIVAR and the VAR. There's a larger discrepancy between FIVAR and VAR when accounting for inflation: Relative to the VAR, the FIVAR attributes about half as much of the variation in inflation to TFP shocks. The inflation series has a strong low-frequency component, coming from the Great Inflation of the 1970s and the Volcker disinflation of the 1980s. The VAR, which cannot generate the same persistence as a FIVAR, leads us to believe that a larger share of inflation can be explained by permanent shifts in TFP. The FIVAR, which provides a better statistical fit at low frequencies, suggests that the movements in inflation are coming more from non-technological sources. The FIVAR also attributes a smaller share of interest-rate movements to TFP shocks, relative to the VAR, but the difference is less pronounced.

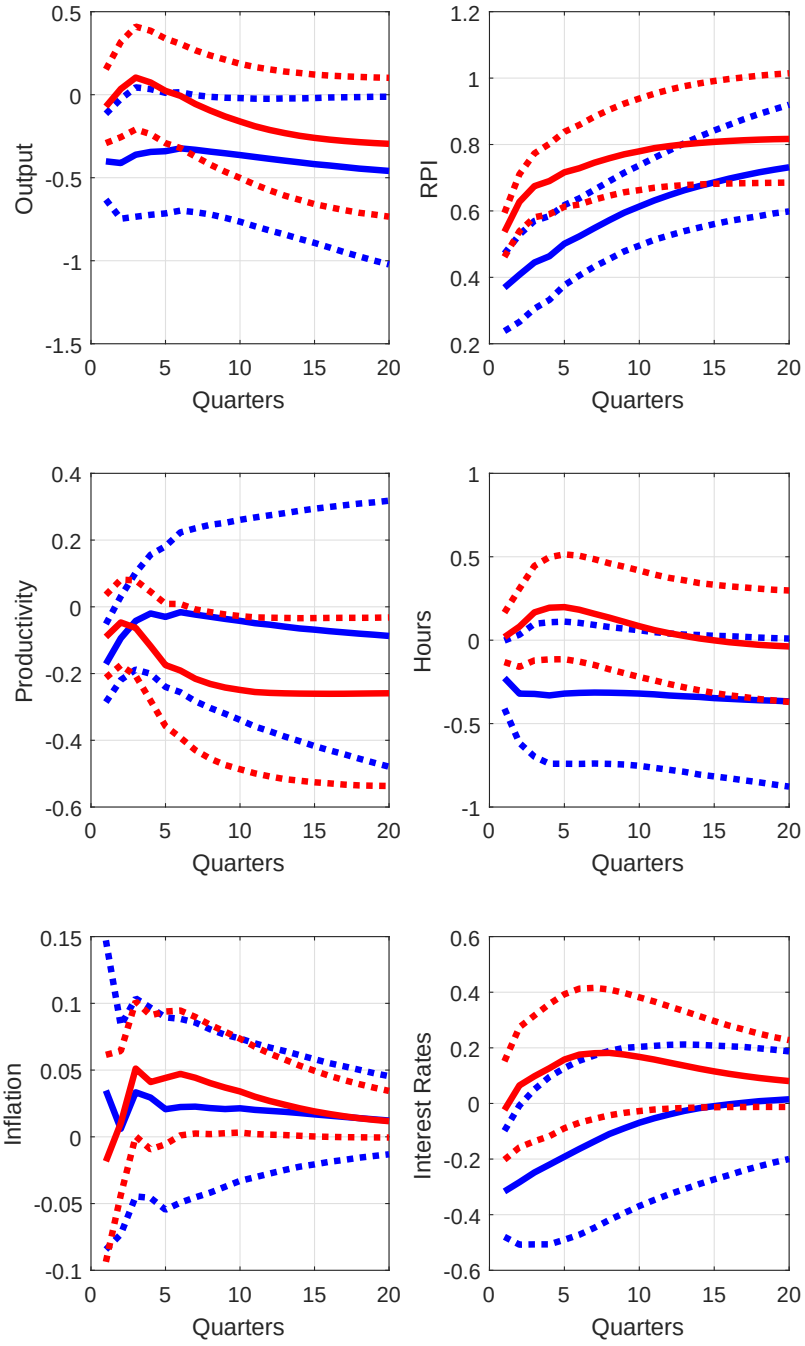
IST Shocks Figure 5 shows the impulse responses from an IST shock, and the FIVAR and the VAR paint substantially different pictures. (The figure depicts a shock that increases RPI, which can be interpreted as a negative movement in IST.) When IST goes down, the VAR basically predicts a once-and-for-all drop in hours, accompanied by relatively little movement in labor productivity, which translates into a once-and-for-all drop in output. In contrast, the FIVAR predicts relatively little movement in hours, accompanied by

Table 3: Variance Decompositions, Full Models

		Model IV: FIVAR			Model IV: VAR		
		Full Spectrum	Business Cycles	Low Frequency	Full Spectrum	Business Cycles	Low Frequency
Output Growth	TFP	9 [5, 15]	4 [1, 11]	21 [9, 36]	9 [5, 17]	5 [2, 15]	21 [8, 39]
	IST	5 [2, 11]	4 [1, 12]	6 [1, 19]	14 [5, 31]	9 [2, 28]	10 [2, 28]
Hours Growth	TFP	22 [11, 36]	21 [9, 36]	9 [3, 21]	23 [9, 41]	22 [7, 41]	12 [3, 27]
	IST	5 [1, 13]	5 [1, 17]	5 [0, 15]	11 [3, 13]	9 [1, 31]	9 [1, 31]
Productivity Growth	TFP	41 [26, 59]	38 [24, 55]	84 [66, 93]	31 [16, 53]	28 [14, 47]	79 [58, 89]
	IST	4 [1, 9]	3 [1, 9]	8 [0, 27]	8 [3, 16]	6 [2, 15]	5 [0, 24]
Inflation	TFP	18 [5, 37]	17 [5, 34]	19 [4, 41]	38 [16, 61]	35 [15, 56]	40 [15, 66]
	IST	8 [2, 27]	8 [2, 19]	8 [0, 33]	8 [2, 35]	6 [1, 26]	9 [0, 40]
Interest Rates	TFP	23 [6, 46]	21 [6, 40]	24 [6, 47]	36 [14, 62]	24 [8, 45]	39 [14, 66]
	IST	7 [0, 31]	6 [1, 20]	8 [0, 32]	12 [3, 41]	16 [5, 43]	10 [2, 41]
RPI Growth	TFP	3 [0, 10]	3 [0, 23]	0 [0, 3]	5 [2, 16]	6 [1, 20]	5 [1, 17]
	IST	76 [58, 88]	73 [52, 87]	95 [88, 98]	40 [22, 61]	32 [13, 57]	73 [54, 85]

Numbers represent the posterior median of the percent contribution of the identified technology shocks to the variance of each variable coming from oscillations in each frequency band. Numbers in square brackets are the 10th and 90th posterior quantiles. "Full spectrum" refers to the unconditional variance; "business cycles" refers to the variance due to oscillations with periods between 1.5 and 8 years; "low frequency" refers to the variance due to oscillations with periods longer than 8 years. All percentages are rounded to the nearest integer.

Figure 5: Impulse Responses from an IST Shock in Model IV



Responses, in levels, from a one-standard-deviation negative shock to IST. The red lines are results from the FIVAR model; the blue lines are results from a VAR without fractional integration. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

a delayed slide in labor productivity, which is mirrored in the path of output. The response of interest rates is weakly positive according to the FIVAR, but negative according to the VAR.

The variance decompositions in Table 3 illustrate the relative (un)importance of IST shocks. Across frequency bands, the FIVAR and the VAR both find that IST shocks account for little of the variation in any of the variables other than RPI. Relative to the VAR, the FIVAR suggests a smaller role for IST shocks in explaining output, hours, productivity, and interest rates, although there's a fair amount of overlap between the FIVAR credible sets and the VAR credible sets. For RPI itself, the FIVAR attributes more much variation to IST shocks than the VAR, especially at business-cycle frequencies. As a point of comparison, several other authors have estimated the contribution of IST shocks to movements in output growth. Using Bayesian DSGE models, Justiniano et al. (2011) and Schmitt-Grohé and Uribe (2012) find that IST shocks account for none of the variance of output.[18] Our results are not that extreme, but the FIVAR in Model IV can only explain 5% of output growth with IST shocks. In contrast, Fisher (2006) uses a (frequentist) VAR to decompose the variance of output growth at business-cycle frequencies; in his baseline specification, he finds that IST accounts for 42% over the period 1955-1979 and 67% over the period 1982-2000. That being said, Fisher uses different data over a different sample period, and Justiniano et al. point out that Fisher may be attributing more importance to IST shocks because his preferred measure of RPI is more strongly countercyclical.[19]

6 Conclusion

FIVARs are under-utilized tools in empirical macroeconomics. In this paper, we have contributed a relatively easy way of expanding the VAR framework to incorporate fractional integration with Bayesian estimation. There are numerous avenues for future research. First, the structural VAR literature has developed various strategies for identifying economic shocks, and most of these approaches can be used with FIVARs as well. In this paper, we analyzed long-run restrictions, and we found differences between the FIVAR's and the VAR's implications for technology shocks. Many cutting-edge identification strategies are explicitly Bayesian, such as the sign restrictions studied by Uhlig (2005), Baumeister and Hamilton (2015), and Arias et al. (2018). Our

[18] Justiniano et al. (2011) decompose the variance of the level of log output at business-cycle frequencies, and Schmitt-Grohé and Uribe (2012) decompose the unconditional variance of the log difference of output. When comparing our results to theirs, note that Schmitt-Grohé and Uribe refer to two different shocks as investment-specific technology: One shock has a temporary effect on the rate at which investment is transformed into installed capital, and the other has a permanent effect on the rate at which consumption goods are transformed into investment. Justiniano et al. refer to the former as a marginal efficiency of investment shock, and the latter as an investment-specific technology shock. Likewise, the shock that we're calling IST corresponds to a non-stationary shock to the rate of transformation between consumption and investment.

[19] Fisher (2006) used a quarterly interpolation of the annual series on the real price of equipment, as estimated by Cummins and Violante (2002). At that time, the NIPA investment deflators did not as thoroughly incorporate quality adjustments, unlike Cummins and Violante. However, Justiniano et al. (2011) point out that the NIPA numbers have subsequently been updated with more quality adjustments in their price indices. Another limitation of the Cummins-Violante data is that it only extends through 2000.

methodology provides a way of applying these insights to situations where the data show signs of fractional integration. Second, it's worth exploring the forecasting properties of Bayesian FIVARs. Giannone et al. (2015), Koop (2013), and Bańbura et al. (2010) investigate the predictive power of Bayesian VARs, with an appropriate amount of shrinkage to accommodate large datasets. It would be interesting to see how the Bayesian FIVAR compares. In particular, the low-frequency features of the FIVAR might be advantageous for forecasting at long horizons.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*(453), 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Arias, J., J. Rubio-Ramirez, and D. Waggoner (2018). Inference based on SVAR identified with sign and zero restrictions: Theory and applications. *Econometrica 86*(2), 685–720.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics 73*(1), 5–59.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Baumeister, C. and J. D. Hamilton (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica 83*(5), 1963–1999.

Bhardwaj, G. and N. R. Swanson (2006). An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics 131*(1-2), 539–578.

Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, 655–673.

Bognanni, M. and E. Herbst (2018). A sequential Monte Carlo approach to inference in multiple-equation Markov-switching models. *Journal of Applied Econometrics 33*(1), 126–140.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2003). What happens after a technology shock? Technical report, National Bureau of Economic Research.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2006). Assessing structural VARs. *NBER Macroeconomics Annual 21*, 1–105.

Cummins, J. G. and G. L. Violante (2002). Investment-specific technical change in the United States (1947–2000): Measurement and macroeconomic consequences. *Review of Economic Dynamics 5*(2), 243–284.

Del Negro, M. and F. Schorfheide (2011). *Bayesian Macroeconometrics*, pp. 293–389. Oxford University Press.

Doppelt, R. and K. O'Hara (2018). Posterior sampling in two classes of multivariate fractionally integrated models. Technical report, New York University.

Fernald, J. G. (2007). Trend breaks, long-run restrictions, and contractionary technology improvements. *Journal of Monetary Economics 54*(8), 2467–2485.

Fisher, J. D. (2006). The dynamic effects of neutral and investment-specific technology shocks. *Journal of Political Economy 114*(3), 413–451.

Francis, N. and V. A. Ramey (2009). Measures of per capita hours and their implications for the technology-hours debate. *Journal of Money, Credit and Banking 41*(6), 1071–1097.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review 89*(1), 249–271.

Gelfand, A. E. and D. K. Dey (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 501–514.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*, 1317–1339.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric reviews 18*(1), 1–73.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Gil-Alana, L. A. and A. Moreno (2009, November). Technology Shocks And Hours Worked: A Fractional Integration Perspective. *Macroeconomic Dynamics 13*(05), 580–604.

Granger, C. W. (1997). On modelling the long run in applied economics. *The Economic Journal 107*(440), 169–177.

Graves, T., R. Gramacy, C. Franzke, and N. Watkins (2015). Efficient Bayesian inference for ARFIMA processes. *Nonlinear Processes in Geophysics Discussions 2*, 573–618.

Henry, M. and P. Zaffaroni (2003). The long range dependence paradigm for macroeconomics and finance. *Theory and applications of long-range dependence*, 417–438.

Herbst, E. and F. Schorfheide (2014). Sequential Monte Carlo sampling for DSGE models. *Journal of Applied Econometrics 29*(7), 1073–1098.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.

Holan, S., T. McElroy, and S. Chakraborty (2009). A Bayesian approach to estimating the long memory parameter. *Bayesian Analysis 4*(1), 159–190.

Hosoya, Y. (1996). The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence. *Journal of Econometrics 73*(1), 217–236.

Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *The Annals of Statistics*, 105–137.

Hsu, N.-J. and F. J. Breidt (2003). Bayesian analysis of fractionally integrated ARMA with additive noise. *Journal of Forecasting 22*(6-7), 491–514.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2011). Investment shocks and the relative price of investment. *Review of Economic Dynamics 14*(1), 102–121.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.

Koop, G., E. Ley, J. Osiewalski, and M. F. Steel (1997). Bayesian analysis of long memory and persistence using ARFIMA models. *Journal of Econometrics 76*(1), 149–169.

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics 28*(2), 177–203.

Liseo, B., D. Marinucci, and L. Petrella (2001). Bayesian semiparametric inference on long-range dependence. *Biometrika 88*(4), 1089–1104.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Lovcha, Y. and A. Perez-Laborda (2015). The hours worked–productivity puzzle: Identification in a fractional integration setting. *Macroeconomic Dynamics 19*(7), 1593–1621.

Pai, J. S. and N. Ravishanker (1998). Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journal of Time Series Analysis 19*(1), 99–112.

Plagborg-Møller, M. (2016). Bayesian inference on structural impulse response functions. Technical report, Harvard University.

Ravishanker, N. and B. K. Ray (1997). Bayesian analysis of vector ARFIMA processes. *Australian Journal of Statistics 39*(3), 295–311.

Ravishanker, N. and B. K. Ray (2002). Bayesian prediction for vector ARFIMA processes. *International Journal of Forecasting 18*(2), 207–214.

Sala, L. (2015). DSGE models in the frequency domains. *Journal of Applied Econometrics 30*(2), 219–240.

Schmitt-Grohé, S. and M. Uribe (2012). What's news in business cycles. *Econometrica 80*(6), 2733–2764.

Sela, R. J. and C. M. Hurvich (2009). Computationally efficient methods for two multivariate fractionally integrated models. *Journal of Time Series Analysis 30*(6), 631–651.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review 97*(3), 586–606.

Sowell, F. (1989a). A decomposition of block Toeplitz matrices with applications to vector time series. Technical report, Carnegie Mellon GSIA.

Sowell, F. (1989b). Maximum likelihood estimation of fractionally integrated time series models. Technical report, Carnegie Mellon GSIA.

Tamaki, K. (2008). The Bernstein-von Mises theorem for stationary processes. *Journal of the Japan Statistical Society 38*(2), 311–323.

Tschernig, R., E. Weber, and R. Weigand (2013). Long-run identification in a fractionally integrated system. *Journal of Business & Economic Statistics 31*(4), 438–450.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics 52*(2), 381–419.

A Proofs

A.1 Proposition 1

The frequency-domain likelihood for the FIVAR process is:

$$\hat{L}_{FIVAR}\left(\mathbf{z}^T \mid \theta\right) = \prod_{k=1}^{K} \pi^{-n} \det\left(\mathbf{f}_{FIVAR}\left(\omega_k \mid \theta\right)\right)^{-1} \exp\left\{-\mathbf{z}_k^* \mathbf{f}_{FIVAR}\left(\omega_k \mid \theta\right)^{-1} \mathbf{z}_k\right\}. \tag{A.1}$$

Notice:

$$\begin{aligned} \det\left(\mathbf{f}_{FIVAR}\left(\omega_k \mid \theta\right)\right)^{-1} &= \det\left(\mathbf{D}\left(\exp\left\{-i\omega_k\right\}\right)^{-1} \mathbf{f}_{VAR}\left(\omega_k \mid \theta\right) \mathbf{D}\left(\exp\left\{i\omega\right\}\right)^{-1}\right)^{-1} \\ &= \det\left(\mathbf{D}\left(\exp\left\{-i\omega_k\right\}\right)\right) \det\left(\mathbf{D}\left(\exp\left\{i\omega_k\right\}\right)\right) \det\left(\mathbf{f}_{VAR}\left(\omega_k \mid \theta\right)\right)^{-1}. \end{aligned} \tag{A.2}$$

Also:

$$\begin{aligned} \mathbf{z}_k^* \mathbf{f}_{FIVAR}\left(\omega_k \mid \theta\right)^{-1} \mathbf{z}_k &= \mathbf{z}_k^* \mathbf{D}\left(\exp\left\{i\omega_k\right\}\right) \mathbf{f}_{VAR}\left(\omega_k \mid \theta\right)^{-1} \mathbf{D}\left(\exp\left\{-i\omega_k\right\}\right) \mathbf{z}_k \\ &= \tilde{\mathbf{z}}_k^* \mathbf{f}_{VAR}\left(\omega_k \mid \theta\right)^{-1} \tilde{\mathbf{z}}_k. \end{aligned} \tag{A.3}$$

Hence:

$$\begin{aligned} \hat{L}_{FIVAR}\left(\mathbf{z}^T \mid \theta\right) &= \left[\prod_{k=1}^{K} \det\left(\mathbf{D}\left(\exp\left\{-i\omega_k\right\}\right)\right) \det\left(\mathbf{D}\left(\exp\left\{i\omega_k\right\}\right)\right)\right] \\ &\quad \times \left[\prod_{k=1}^{K} \pi^{-n} \det\left(\mathbf{f}_{VAR}\left(\omega_k \mid \theta\right)\right)^{-1} \exp\left\{-\tilde{\mathbf{z}}_k^* \mathbf{f}_{VAR}\left(\omega_k \mid \theta\right)^{-1} \tilde{\mathbf{z}}_k\right\}\right]. \end{aligned} \tag{A.4}$$

The second term in square brackets is identical to $\hat{L}_{VAR}\left(\tilde{\mathbf{z}}^T \mid \theta\right)$. To simplify the first term in square brackets, notice that, because $\mathbf{D}\left(\exp\left\{i\omega\right\}\right)$ and $\mathbf{D}\left(\exp\left\{-i\omega\right\}\right)$ are diagonal:

$$\begin{aligned} \det\left(\mathbf{D}\left(\exp\left\{-i\omega\right\}\right)\right) \det\left(\mathbf{D}\left(\exp\left\{i\omega\right\}\right)\right) &= \left[\prod_{j=1}^{n} \left(1 - \exp\left\{-i\omega\right\}\right)^{\delta_j}\right] \left[\prod_{j=1}^{n} \left(1 - \exp\left\{i\omega\right\}\right)^{\delta_j}\right] \\ &= \prod_{j=1}^{n} \left[\left(1 - \exp\left\{-i\omega\right\}\right)\left(1 - \exp\left\{i\omega\right\}\right)\right]^{\delta_j} \\ &= \left[\left(1 - \exp\left\{-i\omega\right\}\right)\left(1 - \exp\left\{i\omega\right\}\right)\right]^{\sum_{j=1}^{n} \delta_j} \\ &= \left[2 - 2\cos\left(\omega\right)\right]^{\sum_{j=1}^{n} \delta_j}. \end{aligned} \tag{A.5}$$

Hence:

$$\begin{aligned}\prod_{k=1}^{K}\det\left(\mathbf{D}\left(\exp\left\{-i\omega_k\right\}\right)\right)\det\left(\mathbf{D}\left(\exp\left\{i\omega_k\right\}\right)\right) &= \prod_{k=1}^{K}\left[2-2\cos\left(\omega_k\right)\right]^{\sum_{j=1}^{n}\delta_j} \\ &= \left(\prod_{k=1}^{K}\left[2-2\cos\left(\omega_k\right)\right]\right)^{\sum_{j=1}^{n}\delta_j} \\ &= \kappa^{\sum_{j=1}^{n}\delta_j}, \end{aligned} \tag{A.6}$$

where $\kappa \equiv \prod_{k=1}^{K}\left[2-2\cos\left(\omega_k\right)\right]$.

A.2 Proposition 2

We will begin by justifying equations (3.17), (3.18), and (3.19). By identity:

$$\tilde{p}\left(\theta \mid \mathbf{x}^T\right) = \tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)\tilde{p}\left(\mathbf{Q}_e \mid \boldsymbol{\delta}, \mathbf{x}^T\right)\tilde{p}\left(\mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta}, \mathbf{x}^T\right). \tag{A.7}$$

For equations (3.17), (3.18), and (3.19) to be true, it must be the case that: $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$ is proportional to (3.17), $\tilde{p}\left(\mathbf{Q}_e \mid \boldsymbol{\delta}, \mathbf{x}^T\right)$ is proportional to the kernel of a Wishart distribution, and $\tilde{p}\left(\mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta}, \mathbf{x}^T\right)$ is proportional to the kernel of a normal distribution. It is therefore sufficient to show:

$$\begin{aligned}\tilde{p}\left(\theta \mid \mathbf{x}^T\right) \propto\ & p\left(\boldsymbol{\delta}\right)\kappa^{\mathbf{1}'\boldsymbol{\delta}}\det\left(\hat{\mathbf{Q}}_a\right)^{-\frac{n}{2}}\det\left(\hat{\mathbf{Q}}_e\right)^{\frac{\nu+T-m}{2}} \\ & \times\det\left(\hat{\mathbf{Q}}_e\right)^{-\frac{\nu+T-m}{2}}\det\left(\mathbf{Q}_e\right)^{\frac{\nu+T-m-n-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\frac{1}{\nu+T-m}\hat{\mathbf{Q}}_e\right)^{-1}\mathbf{Q}_e\right\}\right\} \\ & \times\det\left(\hat{\mathbf{Q}}_a\otimes\mathbf{Q}_e\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\mathbf{a}-\hat{\mathbf{a}}\right)'\left(\hat{\mathbf{Q}}_a\otimes\mathbf{Q}_e\right)\left(\mathbf{a}-\hat{\mathbf{a}}\right)\right\}, \end{aligned} \tag{A.8}$$

where the factor of proportionality abstracts from terms that do not depend on θ.

Let $\bar{\mathbf{a}} \equiv \mathrm{vec}\left(\bar{\mathbf{A}}\right)$. The prior for the VAR parameters $\left(\mathbf{Q}_e, \mathbf{a}\right)$ takes the form:

$$p\left(\mathbf{Q}_e\right) \propto \det\left(\mathbf{Q}_e\right)^{\frac{\nu-n-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\frac{1}{\nu}\bar{\mathbf{Q}}_e\right)^{-1}\mathbf{Q}_e\right\}\right\} \tag{A.9}$$

$$p\left(\mathbf{a} \mid \mathbf{Q}_e\right) \propto \det\left(\mathbf{Q}_a\otimes\mathbf{Q}_e\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\mathbf{a}-\bar{\mathbf{a}}\right)'\left(\mathbf{Q}_a\otimes\mathbf{Q}_e\right)\left(\mathbf{a}-\bar{\mathbf{a}}\right)\right\}. \tag{A.10}$$

We can write the quadratic that appears in $p\left(\mathbf{a} \mid \mathbf{Q}_e\right)$ as:

$$\left(\mathbf{a}-\bar{\mathbf{a}}\right)'\left(\mathbf{Q}_a\otimes\mathbf{Q}_e\right)\left(\mathbf{a}-\bar{\mathbf{a}}\right) = \mathbf{a}'\left(\mathbf{Q}_a\otimes\mathbf{Q}_e\right)\mathbf{a} - 2\mathbf{a}'\,\mathrm{vec}\left(\mathbf{Q}_e\bar{\mathbf{A}}\mathbf{Q}_a\right) + \mathrm{tr}\left\{\bar{\mathbf{A}}\mathbf{Q}_a\bar{\mathbf{A}}'\mathbf{Q}_e\right\}. \tag{A.11}$$

Combining the above with the fact that $\det(\mathbf{Q}_a \otimes \mathbf{Q}_e)^{\frac{1}{2}} = \det(\mathbf{Q}_a)^{\frac{nm}{2}} \det(\mathbf{Q}_e)^{\frac{n}{2}} \propto \det(\mathbf{Q}_e)^{\frac{n}{2}}$ allows us to write the prior over $(\mathbf{Q}_e, \mathbf{a})$ as:

$$\begin{aligned} p(\mathbf{Q}_e)\, p(\mathbf{a} \mid \mathbf{Q}_e) \quad &\propto \quad \det(\mathbf{Q}_e)^{\frac{\nu-n-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\nu\bar{\mathbf{Q}}_e^{-1} + \bar{\mathbf{A}}\mathbf{Q}_a\bar{\mathbf{A}}'\right)\mathbf{Q}_e\right\}\right\} \\ &\quad \times \det(\mathbf{Q}_e)^{\frac{nm}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{a}'\left(\mathbf{Q}_a \otimes \mathbf{Q}_e\right)\mathbf{a} - 2\mathbf{a}'\,\mathrm{vec}\left(\mathbf{Q}_e\bar{\mathbf{A}}\mathbf{Q}_a\right)\right)\right\}. \end{aligned} \tag{A.12}$$

We can write the quadratic form that appears in the VAR likelihood as:

$$\begin{aligned} &\sum_{t=m}^{T}\left(\tilde{\mathbf{x}}_t - \sum_{\ell=1}^{m}\mathbf{A}_\ell\tilde{\mathbf{x}}_{t-\ell}\right)'\mathbf{Q}_e \\ &\qquad \times\left(\tilde{\mathbf{x}}_t - \sum_{\ell=1}^{m}\mathbf{A}_\ell\tilde{\mathbf{x}}_{t-\ell}\right) \quad &= \quad \mathrm{vec}\left(\tilde{\mathbf{Y}}' - \mathbf{A}\tilde{\mathbf{X}}'\right)'\left(\mathbf{I}_{T-p}\otimes\mathbf{Q}_e\right)\mathrm{vec}\left(\tilde{\mathbf{Y}}' - \mathbf{A}\tilde{\mathbf{X}}'\right) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\quad &= \quad \mathrm{tr}\left\{\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\mathbf{Q}_e\right\} - 2\,\mathrm{vec}\left(\mathbf{Q}_e\tilde{\mathbf{Y}}'\tilde{\mathbf{X}}\right)\mathbf{a} + \mathbf{a}'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\otimes\mathbf{Q}_e\right)\mathbf{a}. \end{aligned} \tag{A.13}$$

Writing the quadratic form in this way allows us to write the VAR time-domain likelihood as:

$$\begin{aligned} L_{VAR}\left(\tilde{\mathbf{x}}^T \mid \theta\right) \quad &\propto \quad \det(\mathbf{Q}_e)^{\frac{T-m}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\mathbf{Q}_e\right\}\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\mathbf{a}'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\otimes\mathbf{Q}_e\right)\mathbf{a} - 2\mathbf{a}'\,\mathrm{vec}\left(\mathbf{Q}_e\tilde{\mathbf{Y}}'\tilde{\mathbf{X}}\right)\right]\right\}. \end{aligned} \tag{A.14}$$

Combining the above expressions with some algebra gives us:

$$\begin{aligned} &L_{VAR}\left(\tilde{\mathbf{x}}^T \mid \theta\right) \times \\ &p(\mathbf{Q}_e)\, p(\mathbf{a} \mid \mathbf{Q}_e) \quad \propto \quad \det(\mathbf{Q}_e)^{\frac{\nu+T-m-n-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\nu\bar{\mathbf{Q}}_e^{-1} + \bar{\mathbf{A}}\mathbf{Q}_a\bar{\mathbf{A}}' + \tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\right)\mathbf{Q}_e\right\}\right\} \\ &\qquad\qquad \times \det(\mathbf{Q}_e)^{\frac{nm}{2}} \exp\left\{-\frac{1}{2}\left[\mathbf{a}'\left(\hat{\mathbf{Q}}_a\otimes\mathbf{Q}_e\right)\mathbf{a} - 2\mathbf{a}'\,\mathrm{vec}\left(\mathbf{Q}_e\hat{\mathbf{A}}\hat{\mathbf{Q}}_a\right)\right]\right\}, \end{aligned} \tag{A.15}$$

where we have invoked the definitions of $\hat{\mathbf{A}}$ and $\hat{\mathbf{Q}}_a$. Note that we can complete the square for the quadratic:

$$\mathbf{a}'\left(\hat{\mathbf{Q}}_a\otimes\mathbf{Q}_e\right)\mathbf{a} - 2\mathbf{a}'\,\mathrm{vec}\left(\mathbf{Q}_e\hat{\mathbf{A}}\hat{\mathbf{Q}}_a\right) = \left(\mathbf{a}' - \hat{\mathbf{a}}'\right)\left(\hat{\mathbf{Q}}_a\otimes\mathbf{Q}_e\right)\left(\mathbf{a} - \hat{\mathbf{a}}\right) - \mathrm{tr}\left\{\hat{\mathbf{A}}\hat{\mathbf{Q}}_a\hat{\mathbf{A}}'\mathbf{Q}_e\right\}. \tag{A.16}$$

Also:

$$\det(\mathbf{Q}_e)^{\frac{nm}{2}} = \det\left(\hat{\mathbf{Q}}_a\right)^{-\frac{n}{2}}\det\left(\hat{\mathbf{Q}}_a\right)^{\frac{n}{2}}\det(\mathbf{Q}_e)^{\frac{nm}{2}} = \det\left(\hat{\mathbf{Q}}_a\right)^{-\frac{n}{2}}\det\left(\hat{\mathbf{Q}}_a\otimes\mathbf{Q}_e\right)^{\frac{1}{2}}. \tag{A.17}$$

Hence:

$$
\begin{aligned}
&L_{VAR}\left(\tilde{\mathbf{x}}^T \mid \theta\right) \times \\
&p\left(\mathbf{Q}_e\right) p\left(\mathbf{a} \mid \mathbf{Q}_e\right) \quad \propto \quad \det\left(\hat{\mathbf{Q}}_e\right)^{\frac{\nu+T-m}{2}} \det\left(\hat{\mathbf{Q}}_a\right)^{-\frac{n}{2}} \\
&\qquad \times \det\left(\hat{\mathbf{Q}}_e\right)^{-\frac{\nu+T-m}{2}} \det\left(\mathbf{Q}_e\right)^{\frac{\nu+T-m-n-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\frac{1}{\nu+T-m}\hat{\mathbf{Q}}_e\right)^{-1}\mathbf{Q}_e\right\}\right\} \\
&\qquad \times \det\left(\hat{\mathbf{Q}}_a \otimes \mathbf{Q}_e\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{a}' - \hat{\mathbf{a}}'\right)\left(\hat{\mathbf{Q}}_a \otimes \mathbf{Q}_e\right)\left(\mathbf{a} - \hat{\mathbf{a}}\right)\right\}. \qquad \text{(A.18)}
\end{aligned}
$$

Finally, Proposition 1 allows us to write:

$$
\begin{aligned}
\tilde{p}\left(\theta \mid \mathbf{x}^T\right) \quad &\propto \quad \tilde{L}_{FIVAR}\left(\tilde{\mathbf{x}}^T \mid \theta\right) p\left(\boldsymbol{\delta}\right) p\left(\mathbf{Q}_e\right) p\left(\mathbf{a} \mid \mathbf{Q}_e\right) \\
&= \quad \kappa^{\mathbf{1}'\boldsymbol{\delta}} L_{VAR}\left(\tilde{\mathbf{x}}^T \mid \theta\right) p\left(\boldsymbol{\delta}\right) p\left(\mathbf{Q}_e\right) p\left(\mathbf{a} \mid \mathbf{Q}_e\right). \qquad \text{(A.19)}
\end{aligned}
$$

Combining equations (A.18) and (A.19) yields (A.8), thereby proving the claim.

It remains to characterize the conditional distribution of $\mathbf{Q}_e$, given $\boldsymbol{\delta}$ and $\mathbf{a}$. Notice that the hybrid posterior kernel (A.8) reduces to:

$$
\begin{aligned}
\tilde{p}\left(\theta \mid \mathbf{x}^T\right) \quad \propto \quad & p\left(\boldsymbol{\delta}\right) \kappa^{\mathbf{1}'\boldsymbol{\delta}} \det\left(\mathbf{Q}_e\right)^{\frac{\nu+T+(n-1)m-n-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\nu+T-m\right)\hat{\mathbf{Q}}_e^{-1}\mathbf{Q}_e\right\}\right\} \\
& \times \exp\left\{-\frac{1}{2}\left(\mathbf{a} - \hat{\mathbf{a}}\right)'\left(\hat{\mathbf{Q}}_a \otimes \mathbf{Q}_e\right)\left(\mathbf{a} - \hat{\mathbf{a}}\right)\right\}, \qquad \text{(A.20)}
\end{aligned}
$$

where the above uses the fact that $\det\left(\hat{\mathbf{Q}}_a \otimes \mathbf{Q}_e\right)^{\frac{1}{2}} = \det\left(\hat{\mathbf{Q}}_a\right)^{\frac{n}{2}} \det\left(\mathbf{Q}_e\right)^{\frac{nm}{2}}$. Note that:

$$
\begin{aligned}
\left(\mathbf{a} - \hat{\mathbf{a}}\right)'\left(\hat{\mathbf{Q}}_a \otimes \mathbf{Q}_e\right)\left(\mathbf{a} - \hat{\mathbf{a}}\right) \quad &= \quad \mathrm{tr}\left\{\left(\mathbf{A} - \hat{\mathbf{A}}\right)'\mathbf{Q}_e\left(\mathbf{A} - \hat{\mathbf{A}}\right)\hat{\mathbf{Q}}_a\right\} \\
&= \quad \mathrm{tr}\left\{\left(\mathbf{A} - \hat{\mathbf{A}}\right)\hat{\mathbf{Q}}_a\left(\mathbf{A} - \hat{\mathbf{A}}\right)'\mathbf{Q}_e\right\}. \qquad \text{(A.21)}
\end{aligned}
$$

Hence:

$$
\begin{aligned}
\tilde{p}\left(\theta \mid \mathbf{x}^T\right) \quad \propto \quad & \det\left(\mathbf{Q}_e\right)^{\frac{\nu+T+(n-1)m-n-1}{2}} \\
& \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left[\left(\frac{1}{\nu+T-m}\hat{\mathbf{Q}}_e\right)^{-1} + \left(\mathbf{A} - \hat{\mathbf{A}}\right)\hat{\mathbf{Q}}_a\left(\mathbf{A} - \hat{\mathbf{A}}\right)'\right]\mathbf{Q}_e\right\}\right\}, \qquad \text{(A.22)}
\end{aligned}
$$

where the factor of proportionality abstracts from terms that do not depend on $\mathbf{Q}_e$. The above is the kernel of a Wishart distribution with mean $\tilde{\mathbf{Q}}_e$ and $\nu + T + (n-1)m$ degrees of freedom.

A.3 Proposition 3

Recall that the hybrid posterior kernel is given by equation (A.20). To maximize $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$, we will concentrate out $\mathbf{a}$, given $\boldsymbol{\delta}$ and $\mathbf{Q}_e$; then, we will concentrate out $\mathbf{Q}_e$, given $\boldsymbol{\delta}$. Clearly, from inspecting equation (A.20), we see that, given $\boldsymbol{\delta}$ and $\mathbf{Q}_e$, $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ is maximized by setting $\mathbf{a} = \hat{\mathbf{a}}$. Hence, we seek values of $\boldsymbol{\delta}$ and $\mathbf{Q}_e$ to maximize:

$$p\left(\boldsymbol{\delta}\right)\kappa^{\mathbf{1}'\boldsymbol{\delta}}\det\left(\mathbf{Q}_e\right)^{\frac{\nu+T+(n-1)m-n-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\left(\nu+T-m\right)\hat{\mathbf{Q}}_e^{-1}\mathbf{Q}_e\right\}\right\}. \tag{A.23}$$

Taking a first-order condition with respect to $\mathbf{Q}_e$ and rearranging gives us:

$$\mathbf{Q}_e = \frac{\nu+T-m+\left(m-1\right)n-1}{\nu+T-m}\hat{\mathbf{Q}}_e, \tag{A.24}$$

which maximizes $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$, given a value of $\boldsymbol{\delta}$. (Recall from equation (3.15) that $\hat{\mathbf{Q}}_e$ contains $\boldsymbol{\delta}$.) Plugging the above back into equation (A.23) shows us that maximizing $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ entails choosing $\boldsymbol{\delta}$ to maximize:

$$p\left(\boldsymbol{\delta}\right)\kappa^{\mathbf{1}'\boldsymbol{\delta}}\det\left(\frac{\nu+T-m+\left(m-1\right)n-1}{\nu+T-m}\hat{\mathbf{Q}}_e\right)^{\frac{\nu+T+(n-1)m-n-1}{2}}\exp\left\{-\frac{\nu+T+\left(n-1\right)m-n-1}{2}\right\}, \tag{A.25}$$

which is proportional to the objective function in equation (3.22).

B Additional Results

B.1 Details for Computing the MDD

To implement the modified harmonic mean estimator (4.11), we will construct the density $f(\cdot)$ using standard parametric families that resemble the hybrid posterior. We do not take $f(\cdot)$ to be proportional to the hybrid posterior itself because $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$ is a non-standard distribution: Although we know that this density is proportional to equation (3.17), the normalizing constant is not available in closed form. We will therefore specify $f(\cdot)$ using normal and Wishart densities, while exploiting the structure of the hybrid posterior and truncating the tails, along the lines of Geweke (1999). Let $c(\cdot \mid \psi)$ be the inverse CDF of a chi-squared distribution with ψ degrees of freedom. For $\eta \in (0,1]$, define:

$$\tilde{f}(\theta) \equiv \tilde{f}_\delta(\boldsymbol{\delta})\,\tilde{f}_Q(\mathbf{Q}_e \mid \boldsymbol{\delta})\,\tilde{f}_A(\mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta}) \tag{B.1}$$

$$\tilde{f}_\delta(\boldsymbol{\delta}) \equiv \frac{1}{\eta}\varphi(\boldsymbol{\delta} \mid \boldsymbol{\delta}^\circ, \boldsymbol{\Sigma}_\delta)\,\mathbb{I}\left[(\boldsymbol{\delta}-\boldsymbol{\delta}^\circ)'\boldsymbol{\Sigma}_\delta^{-1}(\boldsymbol{\delta}-\boldsymbol{\delta}^\circ) \leq c(\eta \mid n)\right] \tag{B.2}$$

$$\tilde{f}_Q(\mathbf{Q}_e \mid \boldsymbol{\delta}) \equiv \frac{1}{\eta}\tilde{p}\left(\mathbf{Q}_e \mid \boldsymbol{\delta}, \mathbf{x}^T\right)\mathbb{I}\left[\frac{\mathbf{1}'_{n\times 1}\mathbf{Q}_e\mathbf{1}_{n\times 1}}{\frac{1}{\nu+T-m}\mathbf{1}'_{n\times 1}\hat{\mathbf{Q}}_e\mathbf{1}_{n\times 1}} \leq c(\eta \mid \nu+T-m)\right] \tag{B.3}$$

$$\tilde{f}_A(\mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta}) \equiv \frac{1}{\eta}\tilde{p}\left(\mathbf{a} \mid \boldsymbol{\delta}, \mathbf{x}^T\right)\mathbb{I}\left[(\mathbf{a}-\hat{\mathbf{a}})'\left(\hat{\mathbf{Q}}_a \otimes \mathbf{Q}_e\right)(\mathbf{a}-\hat{\mathbf{a}}) \leq c\left(\eta \mid n^2 m\right)\right], \tag{B.4}$$

where $\varphi(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian density with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, and $\boldsymbol{\delta}^\circ$ and $\boldsymbol{\Sigma}_\delta$ are defined as in equation (4.3). Notice that $\tilde{f}(\theta)$ is a proper probability density, and it emulates the shape of the hybrid posterior: $\tilde{f}_\delta(\boldsymbol{\delta})$ is a truncated Gaussian approximation to $\tilde{p}\left(\boldsymbol{\delta} \mid \mathbf{x}^T\right)$, and $\tilde{f}_Q(\mathbf{Q}_e \mid \boldsymbol{\delta})$ and $\tilde{f}_A(\mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta})$ are truncated versions of $\tilde{p}\left(\mathbf{Q}_e \mid \boldsymbol{\delta}, \mathbf{x}^T\right)$ and $\tilde{p}\left(\mathbf{a} \mid \mathbf{Q}_e, \boldsymbol{\delta}, \mathbf{x}^T\right)$. By construction, $\tilde{f}_\delta$, $\tilde{f}_Q$, and $\tilde{f}_A$ trim the most extreme $100(1-\eta)$ percent of draws from either the multivariate normal distribution or the Wishart distribution. However, these densities are not truncated to the stationary region of the parameter space. Hence, we will specify $f(\cdot)$ as:

$$f(\theta) \equiv \frac{\tilde{f}(\theta)\,\mathbb{I}\left[\boldsymbol{\delta} \in \left(-\frac{1}{2}, \frac{1}{2}\right)^n, \mathbf{a} \in \mathcal{A}\right]}{\mathbb{P}_{\tilde{f}}\left[\boldsymbol{\delta} \in \left(-\frac{1}{2}, \frac{1}{2}\right)^n, \mathbf{a} \in \mathcal{A}\right]}, \tag{B.5}$$

where $\mathcal{A}$ is the subset of $\mathbb{R}^{n^2 m}$ that makes the autoregressive coefficients stationary. We can evaluate the numerator of the above expression easily, but we need to estimate the probability in the denominator. To do so, we can just take a large number of draws from $\tilde{f}(\theta)$, which is straightforward, and compute the probability that the draws are in the stationary region of the parameter space.

In practice, we set $\eta = .95$, although we experiment with several other values to make sure the results aren't too sensitive. Also, to avoid issues with numerical overflow, we evaluate the MDD on a log scale. That is, we define:

$$k(\theta) \equiv L\left(\mathbf{x}^T \mid \theta\right) p(\theta) \tag{B.6}$$

$$K \equiv \max_i \left\{ \log\left(f\left(\theta^{(i)}\right)\right) - \log\left(k\left(\theta^{(i)}\right)\right) \right\}, \tag{B.7}$$

and we implement equation (4.11) by computing:

$$\begin{aligned} \log\left(p\left(\mathbf{x}^T\right)\right) &\approx -\log\left(\frac{1}{N}\sum_{i=1}^{N} \frac{f\left(\theta^{(i)}\right)}{k\left(\theta^{(i)}\right)}\right) \\ &= -K - \log\left(\frac{1}{N}\sum_{i=1}^{N} \exp\left\{\log\left(f\left(\theta^{(i)}\right)\right) - \log\left(k\left(\theta^{(i)}\right)\right) - K\right\}\right). \end{aligned} \tag{B.8}$$

B.2 Hybrid vs. Exact Posterior

Here, we compare results generated using draws from the hybrid posterior and draws from the exact, time-domain posterior. Because θ is high-dimensional, it's not practical to display all the marginal distributions, much less all the contours of the joint distribution. We will therefore focus on two aspects of the posteriors: the log posterior density and the distribution of impulse responses.

Figure 6 compares the log density of the exact, time-domain posterior to the log density of the hybrid posterior for all four models estimated in the paper. Each blue dot corresponds to a draw from the exact posterior, generated by the sequential importance sampler. The black line is the 45-degree line. For Models I-IV, the correlation coefficients between the log density of the exact posterior and the log density of the hybrid posterior are, respectively, .985, .972, .971, and .973. These results suggest that $\tilde{p}\left(\theta \mid \mathbf{x}^T\right)$ provides a reasonable numerical approximation to $p\left(\theta \mid \mathbf{x}^T\right)$ for "typical" values of θ.

Looking at the approximation errors in the impulse response functions serves two purposes, one economic and one statistical. First, by looking at the responses to identified technology shocks, we can see whether the hybrid and exact posteriors lead to comparable inferences about an object of economic substance. Second, although the impulse responses from the non-technology innovations do not have clear economic interpretations, they still have statistical content as a particular normalization of the Wold moving-average representation. That is, the impulse responses are equal to the coefficients of the moving-average representation, so they are sufficient statistics for characterizing the data-generating process. Consequently, we can see whether the hybrid posterior is comparable to the exact posterior for capturing the statistical properties of the data.

For the bivariate models, the hybrid posterior provides an excellent approximation. Figures 7 and 8 show the distributions of impulse responses for Models I and II. ("Shock 1" is the TFP shock, and "shock 2" is an innovation that is orthogonal to the TFP shock.) Besides doing a good job of capturing the statistical

properties of the models, the hybrid posterior generates virtually the exact same beliefs about the effects of TFP shocks as the exact posterior. To summarize the larger models, Figures 9 and 10 show the distributions of impulse responses for Models III and IV. ("Shock 1" is the IST shock, "shock 2" is the TFP shock, and the remaining "shocks" are innovations that are orthogonal to each other and to the identified technology shocks.) Now, with more variables and more parameters, there is a smattering of slight discrepancies between the hybrid and exact posteriors. However, any differences are sufficiently small that an econometrician would draw the same substantive conclusions about the economy's dynamics, regardless of whether she formed her beliefs using the exact posterior or the hybrid approximation to it.

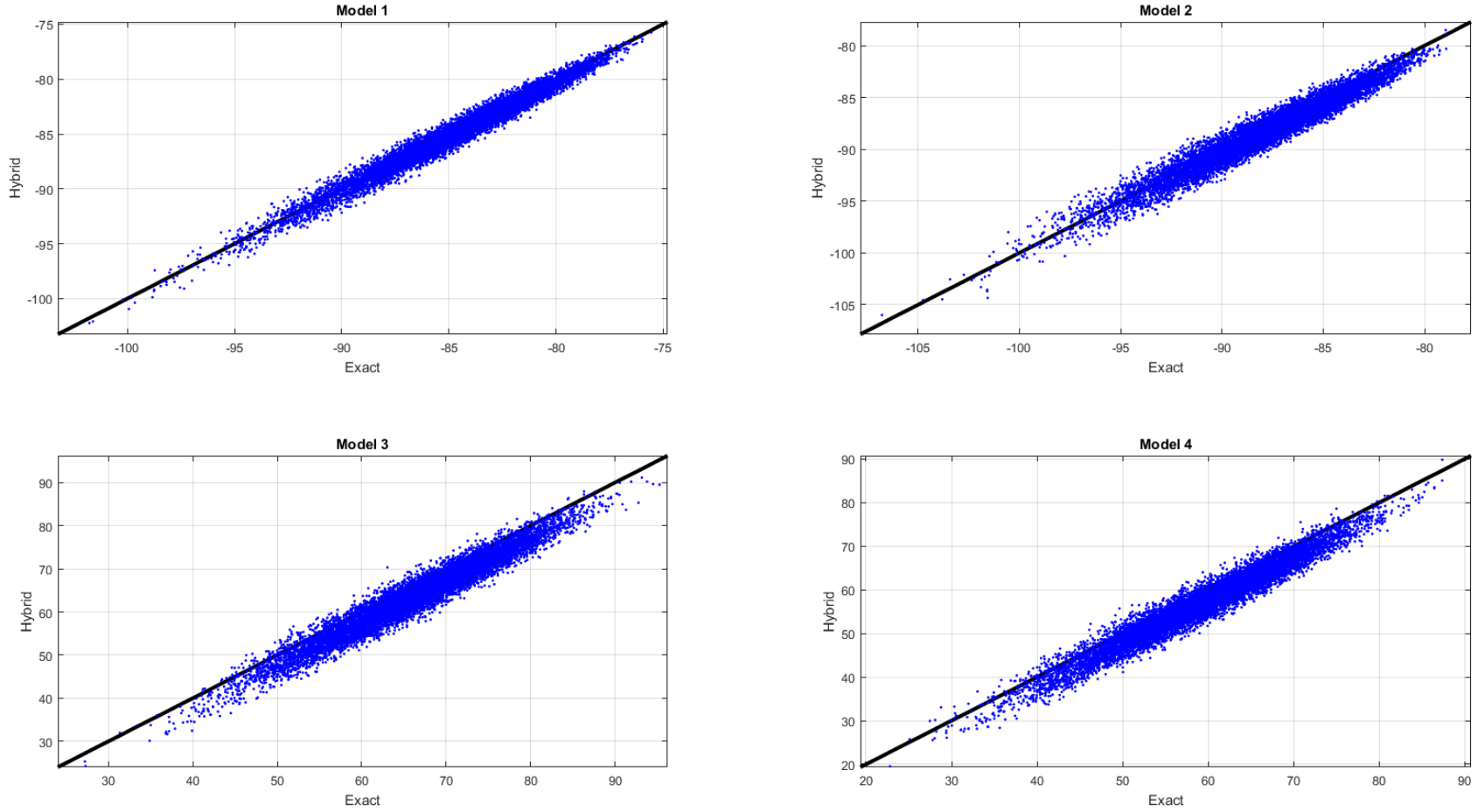
B.3 Results from Model III

As with the bivariate models, estimating the larger models with a FIVAR yields many results that are robust to whether log hours are assumed to be stationary in levels or in differences. Figure 11 shows the impulse responses from a TFP shock when Model III is estimated with a FIVAR (in red) and with a VAR (in blue). The FIVAR results in Figure 11 look similar to those in Figure 4. The posterior median from the VAR shows that hours drop on impact, even when they enter in levels, although the decline is less pronounced than for the VAR in Model IV, which has hours in differences.

Figure 12 shows the impulse responses from a IST shock for Model III. For the FIVARs, the IST responses show more differences between Models III and IV than the TFP responses, but many of the qualitative predictions are the same. The biggest difference is that the FIVAR shows a decline in hours for Model III, but a nearly flat hours response for Model IV. But as with Model IV, Model III shows clear differences between the FIVAR and VAR. Specifically, the VAR predicts a negative response for interest rates, whereas the the FIVAR predicts a weakly positive response. Also, the VAR predicts a stronger decline in hours.

Table 4 shows the variance decompositions for Model III, and the results from the FIVAR are broadly similar to the results from estimating Model IV with a FIVAR. At business-cycle frequencies, the main difference between the FIVAR versions of Model III and Model IV is that the latter attributes more of the variance in RPI growth to IST shocks. Relative to Model IV, Model III exhibits differences between the FIVAR and the VAR that are more pronounced along some dimensions. At business-cycle frequencies, the VAR for Model III suggests that IST shocks are more important for explaining output growth and less important for explaining RPI growth, compared to the VAR for Model IV.

Figure 6: Hybrid vs. Exact Log Posterior Densities



Each dot represents a draw from the exact, time-domain posterior, and the black line is the 45-degree line.

Figure 7: Hybrid vs. Exact Posterior in Models I

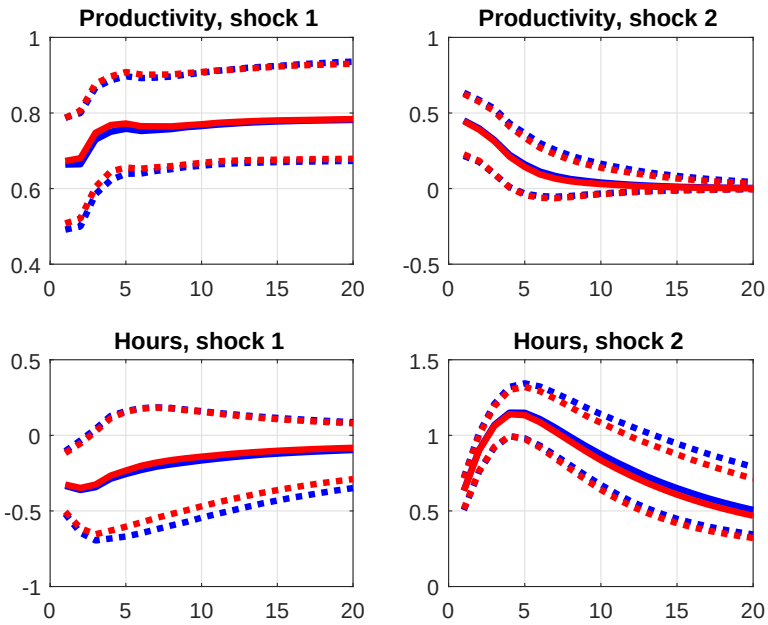
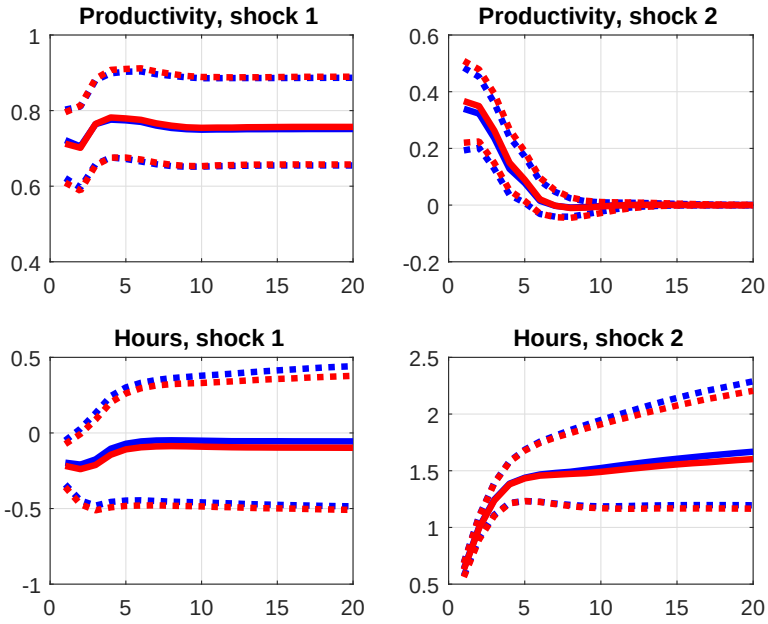
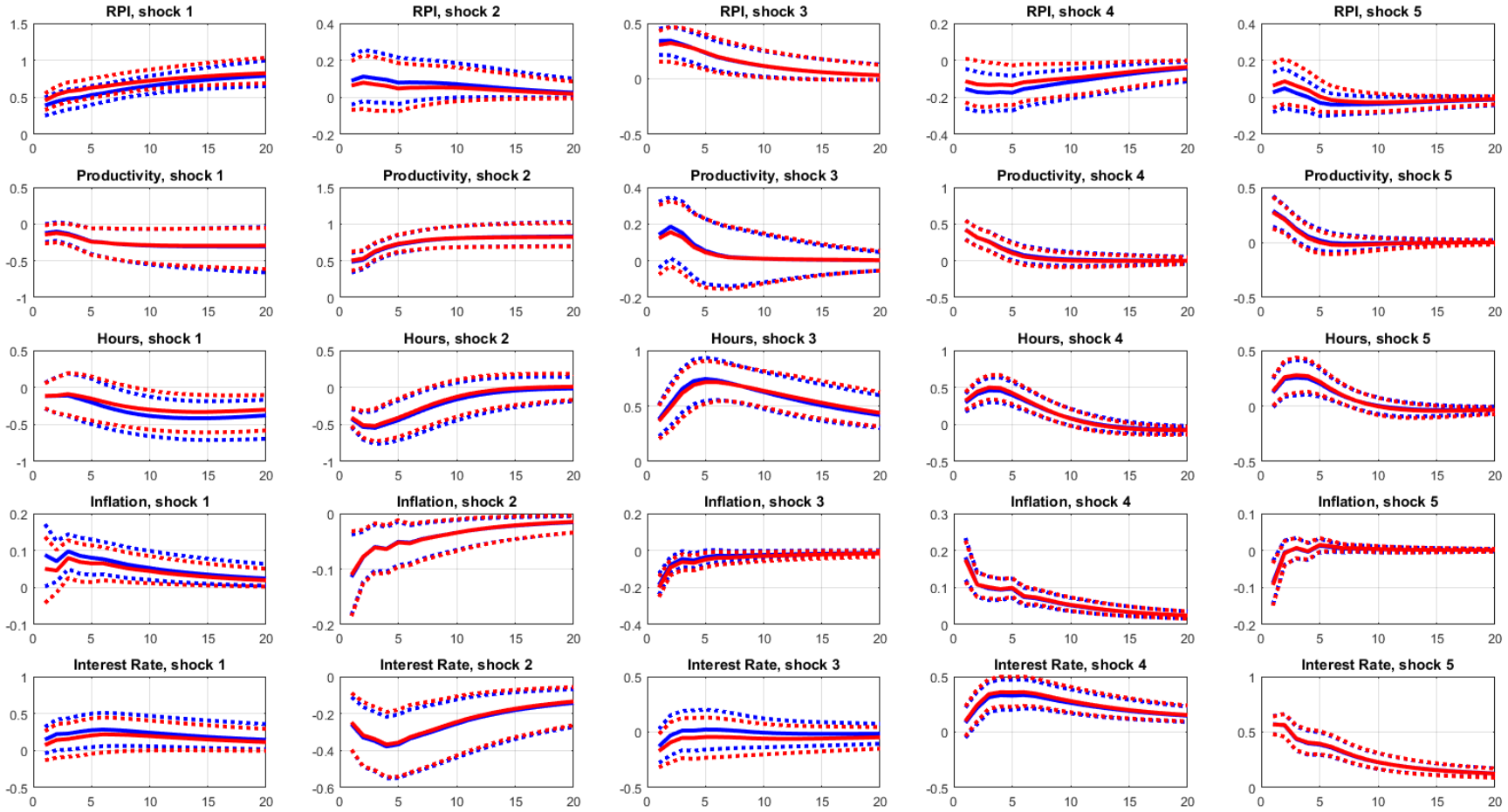


Figure 8: Hybrid vs. Exact Posterior in Models II



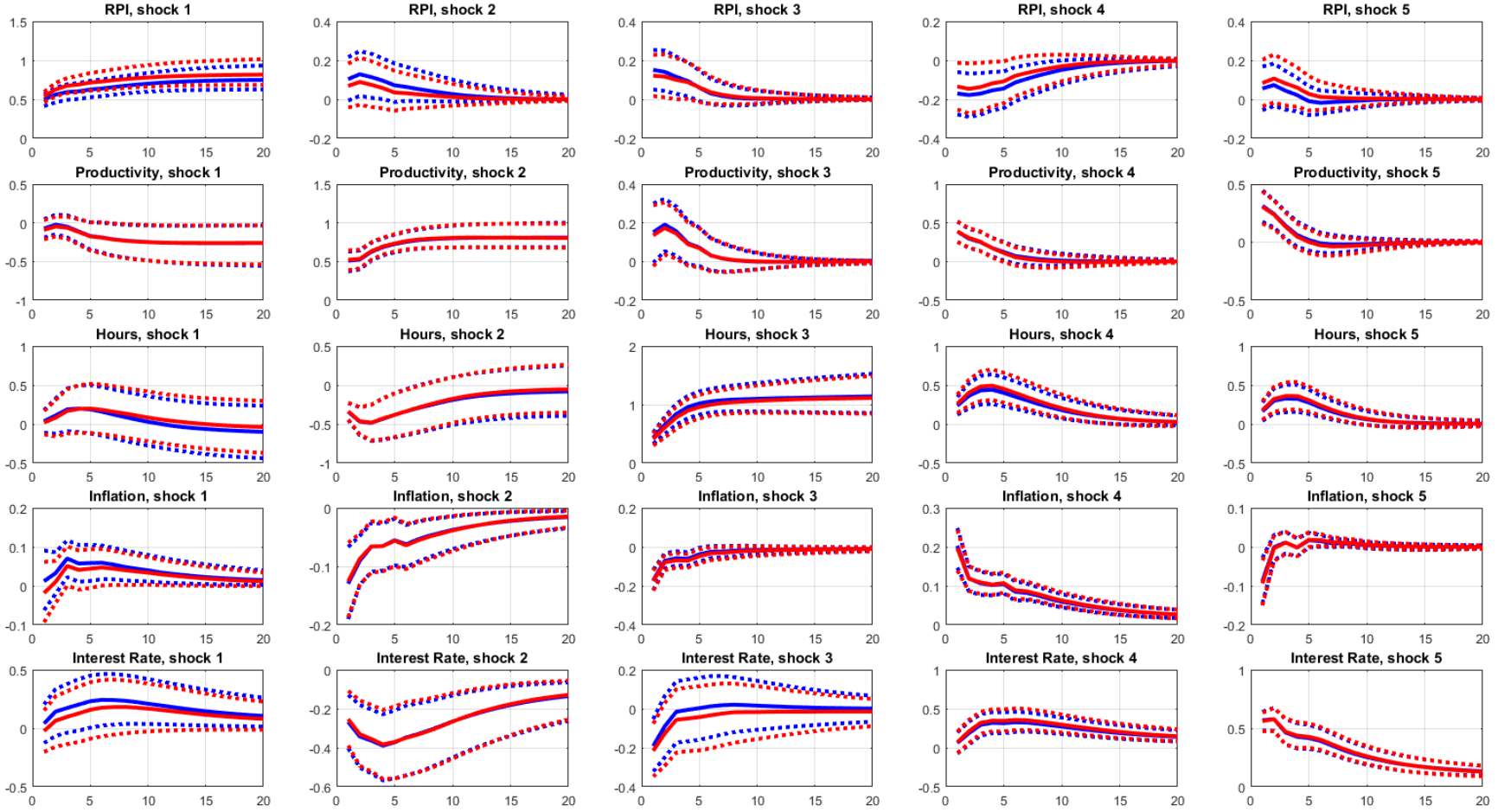
Responses, in levels, from a one-standard-deviation positive shock to TFP in Models I (Figure 7) and II (Figure 8). The red lines represent the exact, time-domain posterior; the blue lines represent the hybrid posterior. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

Figure 9: Hybrid vs. Exact Posterior in Models III



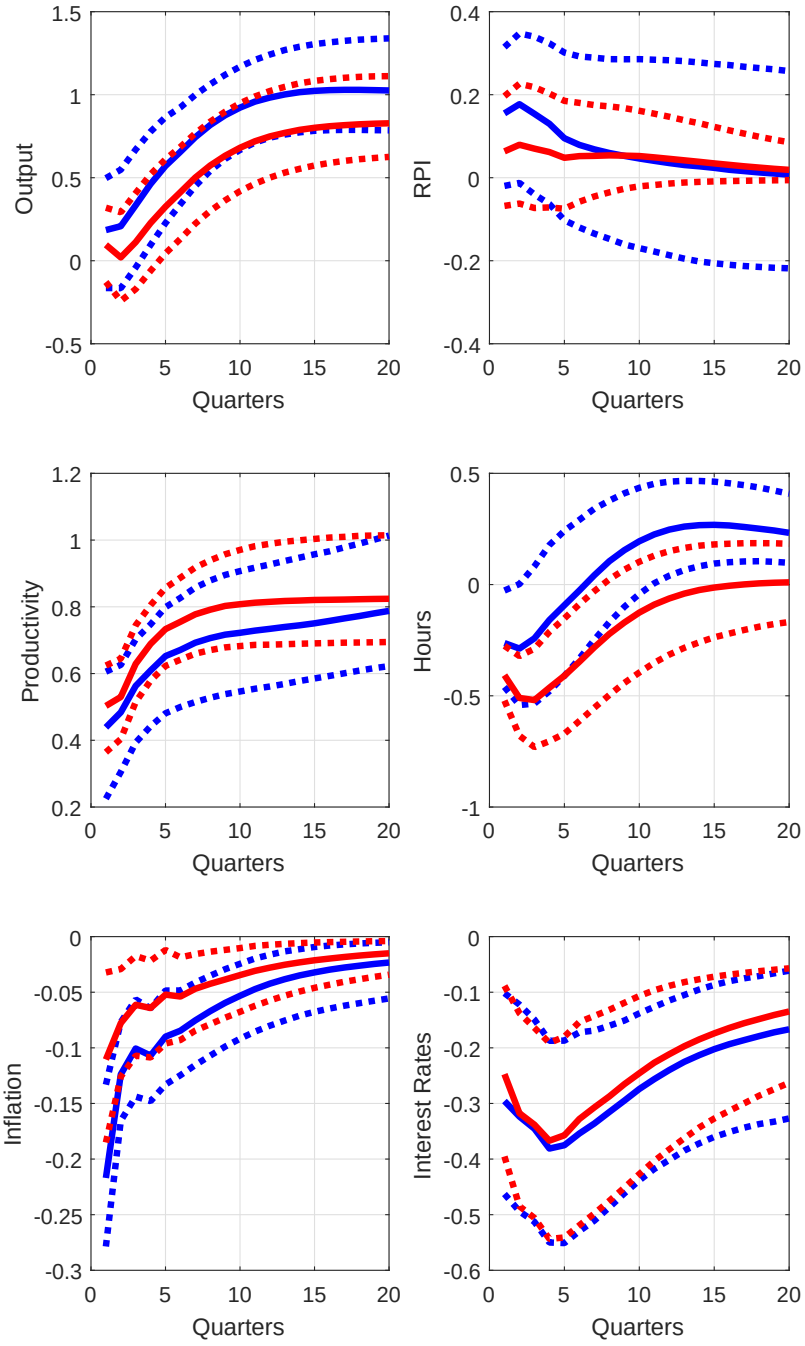
Responses, in levels, from a one-standard-deviation positive shock to TFP. The red lines represent the exact, time-domain posterior; the blue lines represent the hybrid posterior. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

Figure 10: Hybrid vs. Exact Posterior in Models IV



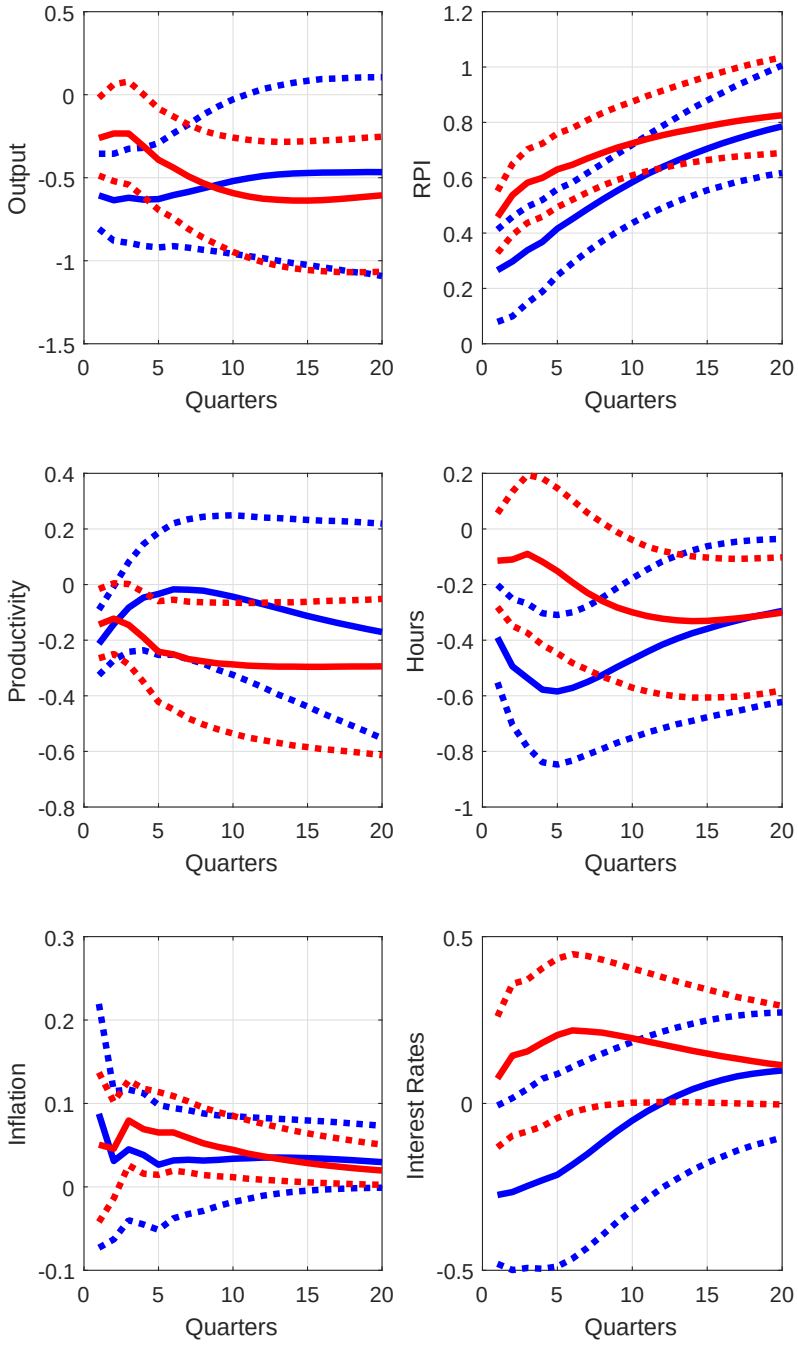
Responses, in levels, from a one-standard-deviation positive shock to TFP. The red lines represent the exact, time-domain posterior; the blue lines represent the hybrid posterior. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

Figure 11: Impulse Responses from a TFP Shock in Model III



Responses, in levels, from a one-standard-deviation positive shock to TFP in Model III. The red lines are results from the FIVAR model; the blue lines are results from a VAR without fractional integration. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

Figure 12: Impulse Responses from an IST Shock in Model III



Responses, in levels, from a one-standard-deviation negative shock to IST in Model III. The red lines are results from the FIVAR model; the blue lines are results from a VAR without fractional integration. Solid lines represent pointwise posterior median responses; dashed lines represent the 10th and 90th posterior quantiles.

Table 4: Variance Decompositions, Full Models

		Model III: FIVAR			Model III: VAR		
		Full Spectrum	Business Cycles	Low Frequency	Full Spectrum	Business Cycles	Low Frequency
Output Growth	TFP	8 [4, 13]	5 [2, 13]	29 [16, 46]	11 [6, 22]	9 [4, 19]	51 [29, 68]
	IST	8 [4, 18]	5 [1, 13]	20 [5, 41]	27 [12, 45]	22 [7, 43]	21 [7, 46]
Hours Growth	TFP	29 [15, 45]	27 [12, 44]	15 [6, 29]	15 [5, 35]	15 [5, 35]	16 [6, 32]
	IST	6 [2, 15]	4 [1, 12]	13 [3, 32]	25 [9, 47]	22 [6, 46]	32 [12, 56]
Productivity Growth	TFP	39 [24, 56]	38 [24, 55]	80 [59, 91]	31 [14, 54]	33 [17, 53]	67 [38, 87]
	IST	5 [2, 12]	4 [1, 10]	11 [1, 31]	11 [5, 21]	9 [3, 18]	10 [2, 39]
Inflation	TFP	14 [2, 34]	14 [2, 34]	15 [2, 37]	38 [14, 65]	42 [19, 64]	37 [11, 67]
	IST	15 [3, 42]	12 [3, 30]	19 [2, 50]	17 [3, 60]	10 [2, 38]	20 [2, 67]
Interest Rates	TFP	21 [5, 44]	20 [5, 39]	21 [5, 45]	31 [12, 56]	22 [5, 45]	33 [11, 59]
	IST	12 [1, 41]	7 [1, 23]	13 [0, 43]	21 [8, 50]	17 [3, 46]	20 [6, 53]
RPI Growth	TFP	3 [0, 11]	3 [0, 13]	1 [0, 6]	9 [2, 28]	11 [2, 33]	5 [0, 18]
	IST	56 [33, 78]	52 [25, 76]	84 [70, 94]	26 [13, 49]	16 [5, 42]	71 [55, 82]

Numbers represent the posterior median of the percent contribution of the identified technology shocks to the variance of each variable coming from oscillations in each frequency band. Numbers in square brackets are the 10th and 90th posterior quantiles. "Full spectrum" refers to the unconditional variance; "business cycles" refers to the variance due to oscillations with periods between 1.5 and 8 years; "low frequency" refers to the variance due to oscillations with periods longer than 8 years. All percentages are rounded to the nearest integer.