Identifying Shocks via Time-Varying Volatility

Daniel J. Lewis*

December 15 2017

The most recent draft can be downloaded here.

Abstract

Under specific parametric assumptions, an $n-$variable structural vector auto-regression (SVAR) can be identified (up to $n!$ shock orderings) via heteroskedasticity of the structural shocks (Rigobon, 2003, Sentana & Fiorentini, 2001). I show that misspecification of the heteroskedasticity process can bias results derived from these identification schemes. I propose a different method that identifies the SVAR up to $n!$ shock orderings using only moment equations implied by a stochastic process for the variance. Unlike previous work, this result requires only weak technical conditions. In particular, it requires neither parametric assumptions nor mutual independence of the shocks. I propose intuitive criteria to select among the orderings and show that this selection does not impact inference asymptotically. As an empirical exercise, I apply this method to Bernanke, Boivin, & Eliasz (2005) and reject their Cholesky assumptions on the structural shocks. However, I confirm their conclusion of monetary non-neutrality and offer several extensions.

Keywords: factor-augmented vector auto-regression, identification, impulse response function, latent variables, monetary non-neutrality, price puzzle, structural shocks, time-varying volatility.

JEL codes: C32, C38, C58, E37, E52.

*Helpful comments provided by Jim Stock, Gabriel Chodorow-Reich, Neil Shephard, Mikkel Plagborg-Møller, Jón Steinsson, Giri Parameswaran, and participants of the Harvard econometrics and macroeconomics lunches.

1 Introduction

A central challenge in structural vector autoregression (SVAR) analysis is how to identify the latent structural shocks that give rise to the observable VAR innovations (one-step ahead reduced-form forecast errors). For example, an innovation to the Federal Funds rate could represent either a true monetary policy shock or the contemporaneous response of monetary policy to changes in macroeconomic conditions. To understand the impact of monetary policy shocks on macroeconomic variables, the shocks must first be isolated from other movements in the data.

In the standard SVAR, the reduced-form innovations, η_t, are expressed as a linear combination of the underlying shocks to the system, ε_t: $\eta_t = H\varepsilon_t$ for some response matrix H. The parameters in these equations are unidentified without further assumptions. Many existing approaches impose assumptions on the response matrix to simplify the problem. These have taken the form of assuming zeros in the short-run (Sims, 1980), zeros in the long-run (Blanchard & Quah, 1986), and sign restrictions (Uhlig, 2005), among others; see Ramey (2016) for a survey. Although progress has been made with these approaches, in many contexts such assumptions are not without controversy.

This paper therefore follows a much smaller literature that considers identification via time-varying heteroskedasticity. Sentana & Fiorentini (2001) and Rigobon (2003) share an important insight. If the variances of the shocks change over time, that variation can identify the parameters of the response matrix of interest. However, their arguments hold only under strict parametric assumptions. Sentana & Fiorentini establish identification conditional on the variance path, which, in practice, means variances must be recoverable from observed data. The method of Rigobon (2003) assumes discrete variance regimes, which must either be determined using external information or estimated. There appears to be no work examining what happens when those parametric assumptions fail to hold. Can their intuition be extended to a general identification argument that does not make use of any such assumptions?

I present an identification approach based on heteroskedasticity that does not depend on any parametric model. I show that if time-varying volatility is present, in any (possibly unspecified) form, identification follows from the autocovariance of the volatility process. Intuitively, working with the autocovariance allows me to abstract from the shocks, which are uncorrelated across periods, focusing instead on the dynamics of the underlying variance process. The presence of such time-varying volatility furnishes equations that identify the response matrix up to a choice of label for each shock, under very general conditions. The strength of this approach lies in the fact that, unlike Sentana & Fiorentini (2001) or Rigobon

(2003), it does not require any information about the path of the variances through time. These results also do not require distributional assumptions or mutual independence of the shocks, unlike e.g. Gouriéroux & Monfort, (2014) and Hyvärinen et al (2010). Frequently, identifying assumptions are a loose approximation of the truth, but the presence of time-varying volatility is uncontroversial in many settings.

In much the same way that Rigobon (2003) observes that a second variance regime doubles the number of equations, I show that a single autocovariance of the reduced-form innovation variances generates a multitude of identifying equations. While Rigobon compares a small number of regimes to yield identification, the autocovariance measures evolution continuously. There is a unique solution when the columns of the volatility process' autocovariance differ adequately so that the equations contain sufficient identifying information.

I outline a wide variety of options for labeling the shocks obtained, which is equivalent to labeling the columns of the response matrix. Indeed, any assumptions that an economist would otherwise employ to identify the matrix itself can analogously be used for the less demanding task of labeling the shock series. Significantly, it is transparent to discuss the impact that such assumptions have on the ultimate estimates of the relevant elements of the response matrix. Since identification relies only on unconditional moments, many more estimation approaches can be used than in the previous literature. Identification via time-varying volatility (TVV-ID) can be implemented directly via Generalized Method of Moments (GMM), without any additional parametric assumptions. A researcher can also make use of any (quasi-)likelihood for the data that implies some form of autocovariance. I compare GMM, likelihood-based methods, and those of Sentana & Fiorentini (2001) and Rigobon (2003) in an extensive simulation study. Likelihood-based approaches perform well, even under misspecification, akin to the results of Quasi-Maximum Likelihood.

TVV-ID works even when heteroskedasticity takes an arbitrary, unknown form. This is most closely compared to Rigobon (2003)'s regime approach when implemented with estimated regimes. I show that when regimes must be estimated, such identification schemes may suffer from substantial bias, which is highly influenced by tuning parameters. Simulation evidence supports this. Further simulations show that TVV-ID, implemented using several estimators, performs favorably for a wide range of data generating processes. Thus, TVV-ID is a more reliable option for researchers not possessing substantial information about the underlying volatility process.

I apply TVV-ID to an important paper in empirical macroeconomics, Bernanke, Boivin, & Eliasz (2005), hereafter BBE, to obtain evidence of monetary non-neutrality. BBE develop the Factor-Augmented Vector Autoregression (FAVAR) methodology, offer evidence of monetary non-neutrality, and make progress in resolving the price puzzle, the surprisingly

ubiquitous finding that inflation responds positively to an interest rate increase. Their paper exploits a Cholesky identification scheme. Using TVV-ID, I strongly reject the assumption of Cholesky structure made by the authors, across estimation approaches. That is, the Federal Funds rate impacts macroeconomic factors contemporaneously. However, most impulse responses remain similar and the conclusions of monetary non-neutrality are robust. For many variables, deviations from the Cholesky structure alter the initial response or amplitude, but wash out fairly quickly. However, the ability of the FAVAR to mitigate the price puzzle is diminished under TVV-ID; it appears that result was largely due to the imposed zeros of the Cholesky assumptions. This aligns well with Ramey's (2016) discussion about features of non-recursive monetary policy estimates.

I also use TVV-ID to extend BBE's FAVAR framework. First, I show that the results are robust to an updated dataset with transformations guaranteeing a higher degree of stationarity in the data. I then consider the possible role of central bank announcement effects, or forward guidance. I find that guidance shocks impact output and inflation in the manner predicted by theory, but that their effects are neither statistically significant, nor economically meaningful. Finally, I consider sub-sample analysis to assess whether the structural relationships underlying the economy were stable over 1960-2001. Despite the dramatic changes of the Volcker period and the Great Moderation, I am unable to reject that the structural relationships remained constant.

In the applied literature, identification via heteroskedasticity has proven very popular. This underscores the importance of understanding the limits of existing identification approaches, as I address with TVV-ID. A glance at recent citations shows that numerous published papers adopt the Rigobon approach alone each year. Prominent examples include Rigobon & Sack (2003, 2004), Craine & Martin (2008), Pavlova & Rigobon (2007), Lanne & Lutkepohl (2008), Ehrmann, Fratzscher, & Rigobon (2011), Eichengreen & Panizza (2016), Ehrmann & Fratzscher (2017), and Hébert & Schreger (2017). Normandin & Phaneuf (2004), Normandin (2004), Doz & Renault (2004), and Lutkepohl & Milunovich (2016), amongst others, have followed path-based identification, mostly in monetary economics and finance. While many applications come from monetary economics and international finance, examples can now be found in public finance (Jahn & Weber, 2016), growth (Islam, Islam & Nguyen, 2017), trade (Lin, Wang, & Weldemicael, 2016, Feenstra & Weinstein, 2017), political economy (Rigobon & Rodrik, 2005, Khalid, 2016), environmental economics (Millimet & Roy, 2016, Gong, Yang, & Zhang, 2017), agriculture and energy (Fernandez-Perez, Frijns, & Tourani-Rad, 2016), education (Hogan & Rigobon, 2009, Klein & Vella, 2009), marketing (Zaefarian et al, 2017), and even fertility studies (Mönkediek & Bras, 2016). Given the possibility of bias in regime-based identification in practice, it is important to understand

the role of less restrictive alternatives like TVV-ID.

The remainder of this paper proceeds as follows. Section 2 describes the identification problem in detail and presents the theoretical results. Section 3 addresses the interpretation of results from TVV-ID. Section 4 discusses estimation strategies. The performance of both identification schemes and estimation approaches is compared in Section 5, based on both theory and simulation studies. The empirical application to BBE follows in Section 6, along with extended analysis. Section 7 concludes.

Notation

The following potentially unfamiliar notation is used in the paper:

$\otimes$	represents the Kronecker product of two matrices
$\odot$	represents the element-wise product of two matrices (i.e. Hadamard product)
$A_{(i)}$	denotes the ith row of matrix A
$A^{(j)}$	denotes the jth column of matrix A
A_{ij}	denotes the ij^{th} element of matrix A
$A^{(-i)}$	denotes all columns of A except for the ith, and similarly for rows and elements
$matdiag\,(A)$	is a vector of the diagonal elements of the square matrix A
$diag\,(a)$	is a diagonal matrix with the vector a on the diagonal
$x_{1:t}$	denotes $\{x_1, x_2, \ldots, x_t\}$
x_{-i}	denotes all elements of x excluding the ith
$E_t\,[\cdot]$	denotes a time-specific expectation, i.e. the mean value of x_t at time t
$\perp\!\!\!\perp$	denotes statistical independence

2 Identification theory

I consider the canonical SVAR setting, relating a vector of innovations, η_t, to unobserved structural shocks, ε_t, by a response matrix, H. η_t is $n \times 1$, obtained from a reduced-form

model, or directly observed. For example, a structural vector auto-regression (SVAR) based on data Y_t would yield $A(L)Y_t = \eta_t$. Similarly, ε_t is $n \times 1$, so H is $n \times n$. Thus,

$$\eta_t = H\varepsilon_t, t = 1, \ldots, T, \tag{1}$$

leaving H and, equivalently, ε_t, to be identified. This section first presents a simple example under special assumptions to build intuition for why this poses an identification problem and why heteroskedasticity may be a useful starting point to solve it. I then develop a representation of higher moments of the reduced-form innovations to serve as identifying equations. The following section establishes conditions under which these equations have a unique solution. I discuss assumptions with respect to H that could be considered restrictive. Finally, I outline in detail how TVV-ID relates to and extends existing identification approaches that exploit heteroskedasticity.

2.1 Intuition for the use of heteroskedasticity

To build intuition, I present standard assumptions underlying Equation (1), and consider how, in this framework, heteroskedasticity can identify H up to $n!$ orderings.

Assumption 0. (temporary) *For all* $t = 1, 2, \ldots, T$,

1. $E_t[\varepsilon_t\varepsilon_t' \mid \sigma_t] = diag(\sigma_t^2) \equiv \Sigma_t$ *(σ_t is the conditional variance of the shocks),*
2. σ_t *is a fourth-order stationary strictly positive stochastic process,*
3. $E[\Sigma_t] = \Sigma_\varepsilon$,
4. *Conditional mean independence of shocks,* $E[\varepsilon_{it} \mid \varepsilon_{-is}] = 0$ *for all* i, *all* $t, s = 1, 2, \ldots T$,
5. H *is time-invariant, invertible, with a unit diagonal normalization.*

Note that the fourth point substitutes conditional mean independence for the usual weaker uncorrelated shocks assumption. While the variance of shocks may change, fixing H means that the economic impact of a unit shock remains the same. It is natural to seek to identify H from the overall covariance of η_t, $E[\eta_t\eta_t'] = \Sigma_\eta$. However, it is well-known that these equations can only identify H up to an orthogonal rotation, Φ (so $\Phi\Phi' = I$). Observe

$$\Sigma_\eta = H\Sigma_\varepsilon H' = (H\Phi)(\Phi'\Sigma_\varepsilon\Phi)(H\Phi)' = H^*\Sigma_\varepsilon^* H^{*'}, \tag{2}$$

where $H^* = H\Phi D_{H,\Phi}$ and $\Sigma_\varepsilon^* = D_{H,\Phi}^{-1}\Phi'\Sigma_\varepsilon\Phi D_{H,\Phi}^{-1}$, with $D_{H,\Phi}$ the matrix that unit-normalizes the diagonal of $H\Phi$. This means that the pairs (H, Σ_ε) and $(H^*, \Sigma_\varepsilon^*)$ are observationally

equivalent. Alternatively, note that due to the symmetry of Σ_η, it offers $n(n+1)/2$ equations, but there are n^2 unknowns. This is the fundamental identification problem posed by the SVAR methodology and indeed many other similar models (e.g. Factor Analysis).

Variation in Σ_t allows the researcher to overcome (2). To build intuition, consider a simple two-variable example, where one structural variance follows a time-varying volatility process and the other takes a fixed value. This admits the simplest form of the Rigobon approach, which yields closed form solutions for H (see e.g. Nakamura & Steinsson, 2016). Without loss of generality, assume σ_{2t}^2 is the variance that changes, while $\sigma_{1t}^2 \equiv \sigma_1^2$, constant. Denote $\sigma_t^2 = \begin{bmatrix} \sigma_1^2 \\ \sigma_{2t}^2 \end{bmatrix}$, and $H = \begin{bmatrix} 1 & H_{12} \\ H_{21} & 1 \end{bmatrix}$. Then the conditional variances of the reduced-form innovations are given by $E_t\left[\eta_t \eta_t' \mid \sigma_t\right] = H \Sigma_t H'$, or

$$\begin{aligned} E_t\left[\eta_{1t}^2 \mid \sigma_t\right] &= \sigma_1^2 + H_{12}^2 \sigma_{2t}^2 \\ E_t\left[\eta_{1t}\eta_{2t} \mid \sigma_t\right] &= H_{12}\sigma_{2t}^2 + H_{21}\sigma_1^2 \\ E_t\left[\eta_{2t}^2 \mid \sigma_t\right] &= \sigma_{2t}^2 + H_{21}^2 \sigma_1^2. \end{aligned}$$

Divide the data into disjoint subsamples, T_A and T_B. In the Rigobon context, these represent high and low volatility regimes. For a random variable x_t, define

$$E_{T_A}\left[x_t\right] \equiv \frac{1}{\#T_A} \sum_{s \in T_A} x_s,$$

the mean over the subsample T_A, where $\#T_A = \sum_{s \in T_A} 1$. Thus,

$$\begin{aligned} E_{T_A}\left[\eta_{1t}^2 \mid \sigma_t\right] &= \sigma_1^2 + H_{12}^2 E_{T_A}\left[\sigma_{2t}^2\right] \\ E_{T_A}\left[\eta_{1t}\eta_{2t} \mid \sigma_t\right] &= H_{12} E_{T_A}\left[\sigma_{2t}^2\right] + H_{21}\sigma_1^2 \\ E_{T_A}\left[\eta_{2t}^2 \mid \sigma_t\right] &= E_{T_A}\left[\sigma_{2t}^2\right] + H_{21}^2 \sigma_1^2, \end{aligned}$$

and similarly for T_B. Now, consider how the expectations change between subsamples. Let

$$\Delta_{|\sigma_t}(\cdot) \equiv E_{T_A}\left[\cdot \mid \sigma_t\right] - E_{T_B}\left[\cdot \mid \sigma_t\right],$$

so

$$\begin{aligned} \Delta_{|\sigma_t}\left(\eta_{1t}^2\right) &= H_{12}^2 \Delta_{|\sigma_t}\left(\sigma_{2t}^2\right) \\ \Delta_{|\sigma_t}\left(\eta_{1t}\eta_{2t}\right) &= H_{12} \Delta_{|\sigma_t}\left(\sigma_{2t}^2\right) \\ \Delta_{|\sigma_t}\left(\eta_{2t}^2\right) &= \Delta_{|\sigma_t}\left(\sigma_{2t}^2\right). \end{aligned}$$

Finally, define

$$\Delta(\cdot) \equiv E\left[\Delta_{|\sigma_t^2}(\cdot)\right],$$

an unconditional expectation over σ_t. Therefore,

$$\begin{aligned}
\Delta\left(\eta_{1t}^2\right) &= H_{12}^2 \Delta\left(\sigma_{2t}^2\right) = E_{T_A}\left[\eta_{1t}^2\right] - E_{T_B}\left[\eta_{1t}^2\right] \\
\Delta\left(\eta_{1t}\eta_{2t}\right) &= H_{12}\Delta\left(\sigma_{2t}^2\right) = E_{T_A}\left[\eta_{1t}\eta_{2t}\right] - E_{T_B}\left[\eta_{1t}\eta_{2t}\right] \\
\Delta\left(\eta_{2t}^2\right) &= \Delta\left(\sigma_{2t}^2\right) = E_{T_A}\left[\eta_{2t}^2\right] - E_{T_B}\left[\eta_{2t}^2\right].
\end{aligned}$$

This provides simple expressions for the difference across subsamples of the unconditional expectation of $\eta_t\eta_t'$. Assuming that $\Delta\left(\sigma_{2t}^2\right) \neq 0$, H_{12} can be identified in closed form:

$$\frac{E_{T_A}\left[\eta_{1t}\eta_{2t}\right] - E_{T_B}\left[\eta_{1t}\eta_{2t}\right]}{E_{T_A}\left[\eta_{2t}^2\right] - E_{T_B}\left[\eta_{2t}^2\right]} = \frac{H_{12}\Delta\left(\sigma_{2t}^2\right)}{\Delta\left(\sigma_{2t}^2\right)} = H_{12}. \tag{3}$$

Stationarity is the only assumption made on the form of the stochastic process for σ_{2t}^2. While the Rigobon identification scheme is motivated by a regime-based process, identification holds even when such a form is a misspecification, provided $\Delta\left(\sigma_{2t}^2\right) \neq 0$, and σ_1^2 is indeed fixed. If there are in fact regimes, they need not be known or correctly specified. However, if the variance process is stationary, the value $\Delta\left(\sigma_{2t}^2\right)$ would be close to zero in population, and identification fails.

Stated this way, Rigobon's approach provides moment conditions based on means of the process, which can yield identification for some processes and not others. However, there is no reason not to consider other moments.

In each period, we can find motivation for an instrumental variables approach. Noting

$$\begin{aligned}
\eta_{2t}\eta_{1t} &= H_{21}\varepsilon_{1t}^2 + H_{12}\varepsilon_{2t}^2 + \varepsilon_{1t}\varepsilon_{2t} + H_{12}H_{21}\varepsilon_{1t}\varepsilon_{2t}, \\
\eta_{2t}^2 &= H_{21}^2\varepsilon_{1t}^2 + 2H_{21}\varepsilon_{1t}\varepsilon_{2t} + \varepsilon_{2t}^2,
\end{aligned}$$

it is clear that H_{12} would be identified if we could obtain the ratio of $H_{12}\varepsilon_{2t}^2$ and ε_{2t}^2. This is not possible as we only observe η_t, and not its separate components. However, we can instrument for ε_{2t}^2 using a lagged value of η_{2t}^2. Note

$$\begin{aligned}
cov\left(\eta_{2t}\eta_{1t}, \eta_{2(t-p)}^2\right) &= H_{12}cov\left(\varepsilon_{2t}^2, \varepsilon_{2(t-p)}^2\right), \\
cov\left(\eta_{2t}^2, \eta_{2(t-p)}^2\right) &= cov\left(\varepsilon_{2t}^2, \varepsilon_{2(t-p)}^2\right),
\end{aligned}$$

by Assumption 0.4 and the fact that σ_1^2 is fixed. H_{12} can then be identified in closed form:

$$\frac{cov\left(\eta_{2t}\eta_{1t}, \eta_{2(t-p)}^2\right)}{cov\left(\eta_{2t}^2, \eta_{2(t-p)}^2\right)} = \frac{H_{12}cov\left(\varepsilon_{2t}^2, \varepsilon_{2(t-p)}^2\right)}{cov\left(\varepsilon_{2t}^2, \varepsilon_{2(t-p)}^2\right)} = H_{12}. \tag{4}$$

This is the familiar IV estimator, where the dependent variable is $\eta_{2t}\eta_{1t}$, endogenous regressor is η_{2t}^2, and the instrument is $\eta_{2(t-p)}^2$. This works because the previous value $\eta_{2(t-p)}^2$ is uncorrelated with all period t terms except those containing ε_{2t}^2. The argument applies for any lag, p. The only assumption on the stochastic process σ_{2t}^2 is that it is stationary and has finite variance, $E_t\left[\varepsilon_{2t}^4\right] < \infty$, for all $t = 1, 2, \ldots, T$. Identification will hold in this case provided

$$cov\left(\varepsilon_{2t}^2, \varepsilon_{2(t-p)}^2\right) \neq 0$$

for some p.

In particular, this requirement that the pth autocovariance of η_{2t}^2 is non-zero is satisfied by a variety of processes for σ_{2t}^2. If the true process is stochastic regime-switching, the condition is met, as there is a non-zero autocovariance at the regime break dates. In a stochastic volatility (SV) process, the condition holds provided the AR coefficients of the variance are non-zero. In a Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) model, provided at least one of the backward-looking parameters is non-zero, it will likewise hold.[1] The condition can be verified for any other stochastic process of interest. This is the crux of TVV-ID: given the structure of the autocovariance of $\eta_t\eta_t'$, comparing elements of the autocovariance (in this simple case, via a ratio) identifies the columns of H.

This flexibility of identification – independent of misspecification – is not shared by the existing approaches. As noted above, if there are not truly variance regimes, or the regimes are not known, using the Rigobon approach runs the risk of a zero denominator. In Sentana & Fiorentini's (2001) work, there is no consideration of whether identification is robust to the imposed GARCH functional form being misspecified. In contrast, the only condition I require is that the stochastic process is stationarity.

Empirical analysis offers motivation for this approach. While it is not possible to consider structural shocks directly, I analyze the autocovariance properties of innovations in an AR(1) process for macro time series. I consider the 128 monthly series spanning 1959-2017 in McCracken & Ng's FRED-MD database. For each set of innovations, I compute the first autocovariance of η_t^2, an analog to the denominator of (4), and test it against the null hypothesis of zero autocovariance. While this testing problem is very noisy, I reject the null

[1] The GARCH model takes the form $\sigma_t^2 = \mu\left(1 - \beta - \Upsilon\right) + \beta\sigma_{t-1}^2 + \Upsilon\varepsilon_{t-1}$, see e.g. Bollerslev (1982).

hypothesis of zero autocovariance for 85 of the series. Figure 1 presents a histogram of the implied AR(1) coefficients of the η_t^2 process. It shows that the distribution is centered well away from zero. I also perform the Variance Stability (VS) test for heteroskedasticity, as described in Dalla, Giraitis, & Phillips (2015). It rejects the null hypothesis of homoskedasticity at the 10% level for 103 of the series, the 5% level for 100, and the 1% level for 93. Thus, it appears that the identifying condition is likely satisfied in much empirical data.

If the denominator is non-zero for multiple p, there are various identifying equations to choose from, (and in principle, the mean restrictions, (3), can be combined with (4)). The identifying moments, (4), can easily be written in the form

$$cov\left(\eta_{2t}^2, \eta_{2t-p}\eta_{1t-p}\right) - H_{12}cov\left(\eta_{2t}^2, \eta_{2t-p}^2\right) = 0$$

and stacked to furnish an overidentified method of moments problem. Alternatively, it might be natural to assume that the variances follow some loose parametric form, like an AR(1), and let this imply the whole series of autocovariances.

In the setting considered above, strong assumptions have been made on dimension and the stochastic process of σ_t^2 ; I now relax those restrictions to yield a more general result.

2.2 Identification via time-varying volatility

Identification via time-varying volatility applies to a much broader class of systems than those outlined to build intuition above. Again, let

$$\eta_t = H\varepsilon_t,\ t = 1, 2, \ldots T.$$

Write $\mathcal{F}_{t-1}$ to denote $\varepsilon_1, \ldots \varepsilon_{t-1}$ and $\sigma_1^2, \ldots \sigma_{t-1}^2$. Dropping Assumption 0, I now adopt Assumption A:

Assumption A. For every $t = 1, 2, \ldots, T$,

1. $E_t\left(\varepsilon_t \mid \sigma_t, \mathcal{F}_{t-1}\right) = 0$ *and* $Var_t\left(\varepsilon_t \mid \sigma_t, \mathcal{F}_{t-1}\right) = \Sigma_t$.
2. $\Sigma_t = diag\left(\sigma_t^2\right), \sigma_t^2 = \sigma_t \odot \sigma_t$.
3. $E_t\left[\sigma_t^2\right] < \infty$.

By explicitly conditioning on σ_t, these assumptions cover both SV and auto-regressive conditional heteroskedasticity-type (ARCH) models (where σ_t is a function of $\varepsilon_1, \ldots \varepsilon_{t-1}$), amongst others.[2]

[2] ARCH takes the same form as GARCH, with the lagged variance term omitted.

Decomposition

I focus on obtaining identifying equations in observable quantities, H, and moments of only the underlying variance process. To do so, I work with a transformation of η_t as my basic data, $\eta_t\eta_t'$. I begin by writing the decomposition,

$$\eta_t\eta_t' = H\Sigma_t H' + V_t,\ V_t = H\left(\varepsilon_t\varepsilon_t' - \Sigma_t\right)H',$$

where σ_t^2 is unknown. Define L to be an elimination matrix, and G a selection matrix (of ones and zeros), see e.g. Magnus & Neudecker (1980).[3] Then

$$\begin{aligned}\zeta_t &= vech\left(\eta_t\eta_t'\right) = vech\left(H\Sigma_t H'\right) + vech\left(V_t\right)\\ &= L\left(H\otimes H\right)vec\left(\Sigma_t\right) + v_t,\ v_t = vech\left(V_t\right) &(5)\\ &= L\left(H\otimes H\right)G\sigma_t^2 + v_t, &(6)\end{aligned}$$

The simplification from (5) to (6) in the first term is surprising and follows due to the diagonality of Σ_t using A.2. This feature plays a key role in properties established later. From the definition of V_t, A.1, and A.3, $E_t\left[V_t \mid \sigma_t, \mathcal{F}_{t-1}\right] = 0$, so $E_t\left[v_t \mid \sigma_t, \mathcal{F}_{t-1}\right] = 0$ and

$$E_t\left[\zeta_t \mid \sigma_t, \mathcal{F}_{t-1}\right] = L\left(H\otimes H\right)G\sigma_t^2.$$

This provides a signal-noise interpretation for the decomposition of the outer product $\eta_t\eta_t'$. It follows from A.3 that I can integrate over Σ_t to obtain $E_t\left[v_t \mid \mathcal{F}_{t-1}\right] = 0$ and similarly that $E_t\left[|v_t|\right] < \infty$. Therefore v_t is a martingale difference sequence.

Properties of ζ_t

Coupled with the decomposition derived above, Assumption B expands on A.3 to allow the establishment of useful properties of $\zeta_t = vech\left(\eta_t\eta_t'\right)$.

Assumption B. For every t,

1. $Var_t\left(\sigma_t^2\right) < \infty$,
2. $Var_t\left(\varepsilon_t\varepsilon_t'\right) < \infty$.

Using these additional assumptions, the autocovariance of ζ_t has a convenient form, given by Proposition 1.

[3] An elimination matrix L is one such that $vech\left(A\right) = Lvec\left(A\right)$. A selection matrix G is one such that $vec\left(ADA'\right) = \left(A\otimes A\right)Gd$ where $d = diag\left(D\right)$.

Proposition 1. *Under Assumptions A.1-2 & B,*

$$Cov\left(\zeta_t, \zeta_s\right) = L\left(H \otimes H\right) G M_{t,s} \left(H \otimes H\right)' L', \ t > s \tag{7}$$

where

$$M_{t,s} = E_t\left[\sigma_t^2 \sigma_s^{2'}\right] G' + E_t\left[\sigma_t^2 vec\left(\varepsilon_s \varepsilon_s' - \Sigma_s\right)'\right] - E_t\left[\sigma_t^2\right] E_t\left[\sigma_s^{2'}\right] G'.$$

This equation has the desired form: it represents an observable quantity, $Cov\left(\zeta_t, \zeta_s\right)$, as a product of H and a matrix composed of moments of the underlying variance process.

Remark. If an additional restriction is imposed on the form of conditional heteroskedasticity, further simplification is possible.

Assumption C. For $t > s$, $E_t\left[\sigma_{it}^2\left(\varepsilon_s \varepsilon_s^{'} - \Sigma_s\right)\right]$ *is diagonal for all* $i = 1, 2, \ldots, n$.

This means that innovations to σ_t^2 cannot depend on any off-diagonal elements of $\varepsilon_s \varepsilon_s^{'}$. Assumption C is trivially satisfied for standard (non-leverage) SV models, which make the stronger assumption that ε_s is independent of σ_t for all s and t. Further, it is satisfied by common GARCH forms (without statistical leverage) found in the literature.[4]

Proposition 2. *Under Assumptions A.1-2, B, & C, (7) simplifies to*

$$Cov\left(\zeta_t, \zeta_s\right) = L\left(H \otimes H\right) G \check{M}_{t,s} G' \left(H \otimes H\right)' L', \tag{8}$$

where $\check{M}_{t,s} = E_t\left[\sigma_t^2 \sigma_s^{2'}\right] + E_t\left[\sigma_t^2 matdiag\left(\varepsilon_s \varepsilon_s' - \Sigma_s\right)'\right] - E_t\left[\sigma_t^2\right] E_t\left[\sigma_s^{2'}\right]$ *, an* $n \times n$ *matrix.*

Proof. See Appendix A.2. □

$\check{M}_{t,s}$ subsequently simplifies further to $Cov\left(\sigma_t^2, \sigma_s^2\right)$ if the process exhibits no conditional heteroskedasticity. Significantly, the imposition of Assumption C reduces the dimension of the nuisance matrix from $n \times n^2$ to $n \times n$. This simpler problem allows slight modifications of the main Theorems below, which are briefly noted.

To summarize, an autocovariance of the vectorization of $\eta_t \eta_t'$, the outer product of the residuals, can be expressed as a product of some known elimination and selection matrices, L and G, the matrix of interest, H, and at most an $n \times n^2$ nuisance matrix, $M_{t,s}$. This is

[4]For use of GARCH models in the SVAR identification literature, see e.g. Normandin & Phaneuf (2004); such work generally restricts the matrices β, Υ to be diagonal. This is actually more restrictive than Assumption C.

remarkably compact for what is essentially a covariance of matrices. Note that at no point is it necessary to assume stationarity, merely that a collection of higher-order moments are finite. All of the expectations used are well-defined for an object at a particular point in time, even if the distribution might be different at another point in time. Since $vech\left(\eta_t\eta_t'\right)$ has dimension $\left(n^2+n\right)/2\times 1$, a single autocovariance yields $\left(n^2+n\right)/2\times\left(n^2+n\right)/2$ equations in $2n^2-n$ unknowns, satisfying the necessary order condition; it remains to show that this system of equations has a unique solution.

Uniqueness

Having derived a set of equations of adequate order to identify H, it remains to show that they yield a unique solution. I make the following assumptions on H:

Assumption D. *H is time-invariant, invertible, with a unit diagonal (without loss of generality).*

Given Assumption D, the conditions under which equation (7) yields a unique solution for H are established by Theorem 1.

Theorem 1. *Under Assumptions A.1-2, B, & D, equation (7) holds. Then H is uniquely determined from (7) (up to labeling of shocks) provided $rank\left(M_{t,s}\right)\geq 2$ and $M_{t,s}$ has no scalar multiple rows.*

Proof. See Appendix A.3. □

Theorem 1 states that (under certain conditions) Equation (7) will yield a unique solution for the relative magnitudes of elements in each column of H. This does not require substantial economic assumptions. The solution is unique up to column order, given the unit-diagonal normalization. However, there are $n!$ column orderings. Thus, while H is meaningfully identified, it may be helpful to think of it as set-identified. While set-identification usually refers to an uncountable set, as in Uhlig (2005), in this case it refers to a small set of identified matrices. This is similar to the sets identified via non-Gaussianity (Gouriéroux & Monfort, 2014), Sims' implementation of subsample identification (see Section 5), or identification in finite mixture models. In some cases, the labeling of shocks is unnecessary (as in many Factor Model settings) – and identification is complete. This is not the case for policy analysis; the labeling issue is discussed in Section 3.1. Compared to existing identification schemes, a key advantage of Theorem 1 and TVV-ID is that it does not presume knowledge of Σ_t, either instantaneously or over periods of time.

In the case where Assumption C holds and $M_{t,s}$ is $n \times n$, the scalar multiple condition is replaced by the following: *"$M_{t,s}$ has no pairs of rows and columns i, j such that both rows i, j are scalar multiples and columns i, j are scalar multiples"*. This slightly weaker condition results from the symmetry of equation (8).

The conditions of Theorem 1 impose interpretable mild restrictions on the process σ_t^2. First, I discuss the rank condition, which is analogous to the requirement in Rigobon identification that the two regimes do not evolve proportionally. In a SV model, the rank assumption requires that the stochastic process σ_t^2 has at least two linearly independent dimensions. For instance, the elements of σ_t^2 cannot all depend linearly on a single common factor and idiosyncratic i.i.d. noise; however, if one element depends linearly and the other quadratically on the factor, the condition is satisfied. Recall that invertibility is assumed – η_t spans the space of structural shocks ε_t. It seems highly unlikely that there truly is only one component to the variances of all macroeconomic shocks to the economy. I turn to an example to illustrate more clearly why the linear independence condition is required for identification.

Example 1. Consider shocks to two closely-related macroeconomic variables. Much of the movement in the variances of each shock is likely driven by a common macroeconomic factor, m_t. Suppose the SV takes the form

$$\sigma_t^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} m_t + \omega_t$$

where ω_t is a 2×1 idiosyncratic component. If there is no persistence in the idiosyncratic components, ω_t, of their variance process, then, assuming stationarity, the autocovariance matrix is given by

$$a \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{9}$$

where a is some scalar. In this case, identification is sought from $L\left(H \otimes H\right) G \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} G' \left(H \otimes H\right)' L'$. This matrix can be re-written as

$$vech\left(a^{1/2} H I_2 H'\right) vech\left(a^{1/2} H I_2 H'\right)' = a \times vech\left(HH'\right) vech\left(HH'\right)'$$

Since, as discussed above, solutions to $vech\left(HH'\right)$ are unique only up to orthogonal rotations, so too are any solutions to the expression on the right-hand-side. Note that similar conclusions follow (with more algebra) if the dimensions of σ_t^2 are related to m_t by different scalars. If however, the second element of σ_t^2 depends on m_t through some arbitrary

continuous function $r(\cdot)$,

$$\sigma_{2t}^2 = r(m_t) + \omega_{2t},$$

if $r(\cdot)$ is non-linear, the autocovariance will not, in general, have the structure in (9) – the precise form depends on the distribution of m_t. In this example, the two shocks could be those to the FFR and a long-term interest rate. The volatility processes are clearly related – the question is the extent to which that relationship is proportional.

In a larger system, beyond, the rank-2 condition, a scalar multiple condition applies to the matrix as a whole. This requirement on the matrix is weaker than a full-rank assumption. Moreover, it is better thought of as a technical assumption pertaining to a pathological case where the linear algebra arguments underlying Theorem 1 break down. In practice, there is little reason to think this condition will precisely fail; rather, it is more likely to lead to a weak identification problem if the condition is close to failing. For a discussion of weak identification in TVV-ID, see Appendix B.1. Nevertheless, in some finance settings, eg. Campbell et al (2017), many volatilities are assumed to move proportionally. If such assumptions are merely approximations to the truth, then weak identification could result. If they are literally true, it is helpful to understand what can still be identified, which motivates the next result.

Even if the scalar multiple condition were to fail, identification is still possible for those columns of H unaffected, as shown by Corollary 1.

Corollary 1. *Under Assumptions A.1-2, B, & D, equation (7) holds. Then $H^{(j)}$ is identified from* (7) *provided rank* $(M_{t,s}) \geq 2$ *and $M_{t,s}$ contains no rows proportional to row j.*

Proof. The result follows from the proof of Theorem 1. □

Again, the symmetric relaxation to *"no scalar multiple rows of row j or no scalar multiple columns of column j"* applies if Assumption C is used.

The dimensionality and scalar multiple assumptions in Theorem 1 can be relaxed further by supplementing additional equations. If, for example, the (often highly informative) mean

$$E_t[\eta_t \eta_t'] = E_t[\zeta_t] \tag{10}$$

is considered, Theorem 1 can be replaced with Theorem 2.

Theorem 2. *Under Assumptions A.1-2, B, & D, equation (7) holds. Then H is uniquely determined from (7) and (10) (up to labeling of shocks) provided* $\begin{bmatrix} M_{t,s} & E_t[\sigma_t^2] \end{bmatrix}$ *has rank* ≥ 2 *and no scalar multiple rows.*

Proof. See Appendix A.5. □

Again, a symmetric extension applies under Assumption C. Theorem 2 requires that, in order for identification to fail, a scalar multiple assumption must also relate $E_t[\sigma_t^2]$ to $M_{t,s}$. Similar arguments can be made, adding in further observable moments, requiring progressively more extensive scalar multiple deficiencies for identification to break down. Corollary 1 can also be extended using the logic of Theorem 2.

As a final theoretical result, I offer a simple corollary, for cases in which the parameters underlying the time-varying volatility are of interest.

Corollary 2. *If the σ_t process is parametrized by θ, and θ is identified from the moments of ε_t, then θ can be identified from moments of η_t if Theorem 1 applies.*

Proof. See Appendix A.6. □

Note that in some cases, additional assumptions may be required to satisfy the identifiability condition of the corollary, for example, normality of the disturbances ε_t. A brief discussion is offered following the proof in the Appendix.

Overidentification and Assumption D

Even for $n = 2$, the system of equations is overidentified, with the degree of overidentification increasing in n; this means that tests of the model exploiting overidentification, like a $J-$ test, can be conducted. This is an advantage over many identification approaches in this setting, where strong assumptions are required to yield a just-identified model, making specification tests rare. The meaningful modeling assumptions made are that H is invertible and fixed throughout time. A growing literature considers issues surrounding the invertibility of H, (Stock & Watson, 2017, Plagborg-Møller, 2017). In short, if there are more than n underlying shocks in the economy, the true H is non-invertible. However, it is almost always necessary to assume H is invertible for identification purposes. Thus, a test indicating misspecification likely relates to the invertibility assumption.

The other substantive assumption made is that while TVV-ID focuses on the instability of the variances of structural shocks, Assumption D requires that H remain fixed. While this may seem inconsistent, there are several points to consider. First, to the best of my knowledge, no existing identification scheme successfully handles time-varying H. Even simple Cholesky identification, when the true structure is Cholesky, does not identify the mean of H if H is time-varying. Compared to other schemes that assume time-varying volatility, such as Rigobon (2003), TVV-ID is in a better position to consider sub-sample estimation to evaluate the stability of H over time. Rigobon's identification argument fails when H is not

the same across sub-samples. Further, since Rigobon is already working with sub-samples, it becomes difficult to further subdivide in order to isolate both the variance regimes and separate H regimes (which are likely related). Allowing H to vary presents an interesting econometric problem, which is a prominent part of an ongoing research agenda. However, even if H varies, provided it does so at a slower rate than the variances, identification can still hold; H will be locally stationary over intervals over which the variances are not.

There is also good reason to believe that the policy transmission mechanisms captured by H do indeed move more slowly than the variances in the economy. This notion appears in theoretical work; for example, Barro & Liao (2017) split volatility into short-run and long-run components, which move around more slowly. If agents in the economy mainly respond to long-run movements (due to adjustment costs, rational inattention, etc.), then H will also be slow-moving. In addition, in my empirical study I am unable to reject the null hypothesis that H was stable in the pre-Volcker period (1960-1984) and the following Great Moderation, despite substantial perceived changes in macroeconomic dynamics. Regardless, should a researcher remain worried about the assumption of a fixed H, it is natural to apply a $J-$test of the underlying model. Further, Andrews (1993) develops tests for parameter instability in a GMM context, for example the sup-Wald test, the conditions for which are satisfied for a variety of time-varying volatility models.[5]

Stationarity

Note that at no point is stationarity assumed for the process σ_t^2; I rely only on the existence of necessary moments. This is possible because, conditional on some distribution over outcomes at an initial point, it is natural to form moments over future values of a process, σ_t^2. The moments need not be the same – $E_t\left[\sigma_t^2\right] \neq E_s\left[\sigma_s^2\right]$ for $t \neq s$, but both can be well-defined. If $Cov\left(\zeta_t, \zeta_s\right)$ is known for any t, s pair, the identification argument holds. However, stationarity can play a role in estimation, where assumptions are required for $Cov\left(\zeta_t, \zeta_s\right)$ to be well-estimated.

Connection to signal processing TVV-ID has important connections to the signal processing literature. There is a duality with the problem of recovering the signal, for example

[5]The use of GMM to estimate TVV-ID models is discussed below. The less-familiar assumptions needed in Andrews (1993), those of Near-Epoch Dependence (NED), can be replaced by stronger properties that hold for both GARCH and SV processes. Lindner (2009) shows that GARCH satisfies β-mixing (and thus α-mixing with exponential rate) and Davis & Mikosch (2009) show that SV models inherit the mixing properties of the log-variance process. Andrews' (1983) results show that an AR(1) variance process is $\alpha-$mixing with exponential rate. These mixing properties can be shown to imply NED; see Davidson (1994) Chapter 17 for additional background.

the volatility of ε_t, from a noisy measurement, η_t. In fact, this problem is studied by electrical engineers, in particular in relation to medical devices such as electroencephalograms, (see Blanco & Mulgrew, 2005) or by geophysicists, in relation to earthquake detection, as discussed in Bharadwaj, Demanet, & Fournier (2017). As a signal extraction problem, Blanco & Mulgrew (2005) and Blanco et al (2007) have resorted to higher moments, in a framework imposing independence of the noise across dimensions of the measurement, based on Multivariate Independent Components Analysis (MICA). Of course, while that assumption may be plausible in some contexts, it is not in macroeconomics, where the noise is inherently part of the shocks themselves that are mixed to form the measurement (the reduced-form innovation), as opposed to the noise impacting each dimension of the already-mixed signal independently. It is for this reason that, while these authors rely on contemporaneous fourth moments or cumulants, i.e. $E\left[vec\left(\eta_t\eta_t'\right)vec\left(\eta_t\eta_t'\right)'\right]$, I make use of lagged fourth moments i.e. $E\left[vec\left(\eta_t\eta_t'\right)vec\left(\eta_{t-1}\eta_{t-1}'\right)'\right]$.

2.3 Extending the existing literature

Since TVV-ID requires less restrictive assumptions, it holds in virtually any case where previously developed identification schemes apply. While Sentana & Fiorentini (2001) show that the presence of time-varying volatility is sufficient to identify this model, conditional on the path of variances, TVV-ID demonstrates that knowing the values the variance takes is not necessary for identification. In particular, they prove that H is identified from the path $H\Sigma_t H'$ up to column order provided the stochastic processes in σ_t^2 are linearly independent and $T \geq 2$. However, Sentana & Fiorentini's ability to apply this result is restricted by its reliance on the path of $H\Sigma_t H'$ for $t = 1, ..., T$. In most applications, it is only possible to estimate the noisy $\eta_t\eta_t' = H\varepsilon_t\varepsilon_t'H'$; $H\Sigma_t H'$ is never observed directly. Thus, to apply their result, it is necessary to propose some method that provides a one-to-one mapping between (η_t, H) and $\sigma_{1:T}^2$ (this also implies the path $\varepsilon_{1:T}$). Then, a matrix H can be chosen that satisfies some criterion – such as maximizing the joint likelihood of $\sigma_{1:T}^2$ and $\varepsilon_{1:T}^2$. Sentana & Fiorentini propose the only likelihood-based choice discussed in the literature, assuming a GARCH structure for σ_t^2; that is, σ_t^2 evolves depending only on its past values and past values of $\varepsilon_t\varepsilon_t'$. Thus, σ_t^2 is entirely predictable based on η_{t-1} and H, satisfying the requirement of a one-to-one mapping.[6] It is concerning that our ability to exploit an identification argument is dependent on a functional form assumption, particularly in applications of identification via heteroskedasticity removed from finance, where the GARCH structure is most familiar. While estimates may often be sensitive to functional form assumptions, here, identification

[6] Milunovich & Yang (2013) offer an alternative proof of identification under the GARCH assumptions based on the Jacobian of the moment equations.

itself is literally dependent on such an assumption, or some other restrictive method to map (η_t, H) to $\sigma^2_{1:T}$ one-to-one. TVV-ID clearly nests the GARCH-based estimation of Sentana & Fiorentini (2001) – that functional form implies a matrix $M_{t,t-1} \equiv M_1$ based on the first autocovariance of the stationary variance process, so Theorem 1 can be applied.

Rigobon (2003) simplifies the insight of Sentana & Fiorentini, showing that two or more variance regimes are sufficient to identify H. While $H\Sigma_t H'$ cannot be observed from the data, it is essentially possible to observe $H\Sigma_A H' = HE\left[\Sigma_t \mid t \in A\right] H'$, the mean over the sub-sample $A \subset T$, by LLN, provided the set A is large and σ^2_t is stationary within the sub-sample. If there is a second such set, B, and the rows of $\begin{bmatrix} diag\left(\Sigma_A\right) & diag\left(\Sigma_B\right) \end{bmatrix}$ are not proportional, H is identified up to column order, a special case of Sentana & Fiorentini's Proposition 3. Intuitively, with one regime (the whole sample) there are $\left(n^2+n\right)/2$ equations in n^2 unknowns. Adding a second regime yields twice as many equations, n^2+n, with only an additional n parameters, leaving a total of n^2+n unknowns. The requirement of linear independence ensures the rank condition holds. Additional regimes offer overidentification. As a matter of revealed preference, this approach has been most popular in the identification via heteroskedasticity literature. TVV-ID will work in almost all cases where a regime-based approach could be used. Switching between regimes is a parametric form of time-varying volatility, and yields a matrix of the form $M_{t,s}$, averaging over regimes and breaks. However, $M_{t,s}$ will only satisfy the technical conditions when a break occurs between s and t. Thus, if identification is attempted using overall sample moments where the number of breaks is much less than T, identification will fail asymptotically.

When there are two clear variance regimes, Rigobon's scheme is compelling; it is more difficult when variance regimes must be imposed or estimated. The same is true when it seems more plausible that variances just fluctuate continuously. External information can convincingly isolate periods of high and low volatility, as in the original Rigobon (2003) paper, Rigobon & Sack (2004), and Lanne & Lutkepohl (2008), up to more recent papers such as Nakamura & Steinsson (2016). These studies make arguments such as "the volatility of the monetary policy shock will be higher on monetary announcement days" or "the historical record indicates periods of crisis and thus high volatility in Latin American currency markets". Splitting the sample on such a basis furnishes the two subsamples required for identification. When such information is not available, or the variance is thought to change continuously, regimes must be imposed or estimated, using some sort of threshold rule. Such estimation is most analogous to the spirit of TVV-ID – assuming the presence of heteroskedasticity, and seeking to identify H without imposing additional information. Examples include Rigobon & Sack (2003), Pavlova & Rigobon (2007), and Ehrmann, Fratzscher, & Rigobon (2011). The difficulties resulting from the estimation of regimes, which will in-

herently be endogenous, including the possibility of substantial bias, are discussed in Section 5. The presence of such bias in the Rigobon methodology in contexts featuring arbitrary heteroskedasticity of unknown form serves to further set TVV-ID apart.

3 Interpretation of results

Having identified H through TVV-ID, there are myriad approaches to labeling the resulting structural shocks, or, equivalently, the columns of H. This is a necessary step for policy analysis. In this section, I first discuss a range of labeling approaches that might be appropriate in various contexts. Second, I offer results that show that if the labeling procedure has certain asymptotic properties, it does not impact inference on H. Finally, I discuss practical ways in which the researchers can report their results transparently and accentuate the robustness of their findings.

3.1 Labeling of shocks

Labeling can be viewed as part of the identification problem, as it is still necessary to shrink the set of candidate H matrices to obtain a point estimate. The same problem arises in the identification schemes proposed by Gouriéroux & Monfort (2014) and Lanne, Meitz, & Saikonnen (2017) who discuss identification via non-Gaussianity. Gouriéroux & Monfort (2014) do not address the issue, and Lanne, Meitz, & Saikonnen (2017), offer a purely statistical method of selecting the labeling, based on a pre-defined normalization of H and an arbitrary preference for the relative of magnitude of subsequent elements in each column. Ludvigson & Ng (2016) discuss "winnowing constraints" to eliminate possible solutions for H; loosely speaking, these comprise both event constraints (certain shock series must take values above/below certain thresholds during important periods) and correlation constraints between external variables and the structural shocks. As such, in considering candidate approaches, it is important to emphasize that while some will yield a unique identified shock series for all datasets, some will only do so in certain cases, otherwise simply ruling out a number of impermissible shock series, leaving some observationally equivalent. In general, the latter approaches simply apply a weaker form of the assumptions found in the former. It is notable that for virtually any macroeconomic identification scheme one could employ to identify a latent variables model, there is a weaker analog that can be used here to choose between the shock series, once the series have been identified via time-varying volatility. The present discussion is not meant to offer an exhaustive account of labeling methods available; rather, it is intended to be suggestive of the variety of assumptions that may be adopted.

In reality, each applied setting will lend itself to its own particular set of assumptions, and a researcher ought to choose carefully based on the data to be considered, as she would otherwise be forced to do before identifying H itself in the first place.

First, consider assumptions that map at most one shock to a label. Note that if a researcher intends to label all shocks, not just a single policy shock of interest, some of these assumptions may map one shock to multiple labels. They are applied post-estimation, with the exception of those restricting elements of the H matrix, which must be imposed in the estimation routine. Examples are framed in the setting of a standard three-variable monetary policy VAR.

1. Stock (2008) introduces the "external instruments" framework. He shows that, for an instrument Z_t, if $E\left[Z_t \varepsilon_{jt}\right] = 0$ for $j \neq i$, and non-zero for $j = i$, the ith column of H is identified. To label the shock series, rather than Stock's sharp exogeneity assumption, it is possible to use the weaker assumption that Z_t is better predicted by the shock series of interest than any of the others. Thus, for the simple regression $Z_t = \beta_i \varepsilon_{it} + \nu_t$, assume $\frac{var\left(\hat{Z}_t^i\right)}{var(Z_t)} > \frac{var\left(\hat{Z}_t^j\right)}{var(Z_t)}$ where $\hat{Z}_t^i = \hat{\beta}_i \varepsilon_{it}$, or vice versa. For example, "the monetary policy shock should better explain innovations to the price of interest rate futures than the unemployment or inflation shock."

2. Similar to external instruments, zeroth (or higher) order forecast error variance decomposition (FEVD) can be used if a researcher thinks that the shock of interest is a stronger driver of an *internal* regressor than any of the others. For example, "if monetary policy decisions are rarely dominated by *simultaneous* movements in macro variables, a greater share of variation in the residual of the interest rate series should be predicted by the monetary policy shock than the unemployment or inflation shocks at a contemporaneous horizon."

3. In a Rigobon sub-sample setting, or indeed that considered by Sims (2014), it is generally assumed that a certain series exhibits a larger variance change from the low-volatility sub-sample to the high-volatility sub-sample. Similarly, a researcher can rank various moments of the shock series volatilities. It is important to first normalize these series, as their scale will vary depending on the ordering (and thus normalization) in H. For example, "the volatility process of the monetary policy shock has higher variance than the unemployment or inflation shocks due to slowly changing structural factors impacting the latter variables."

4. Restricting one or more elements of the column of interest, $H^{(i)}$, yields identification as it is then clear that the shock series corresponding to the restricted column of the matrix

is the relevant one. For example, "the contemporaneous response of the unemployment rate to the monetary policy shock is zero (or has some fixed relationship to other coefficients)." Note that while this carries the flavour of Cholesky decomposition, it is weaker, as it requires only one assumption on the column of interest.

5. Conversely, imposing assumptions on all other columns of H, denoted $H^{(-i)}$, means that the shock corresponding to the unrestricted column is that of interest. For example, "apply a partial-Cholesky decomposition imposing no relative ordering between unemployment and inflation but ordering the interest rate last." This is similar to the "Slow-R-Fast" scheme.

If such assumptions are deemed too stringent, a weaker class of assumptions may yield a mapping of multiple shocks to a single label. Assumptions thus limiting the identified set include the following:

1. In a weaker version of 3, a partial ordering of higher moments can perform a similar role. For example, "the variance of the unemployment shock should be lower than that of the monetary policy shock," or "the volatility of the unemployment shock should have lower variance over the business cycle than that of the interest rate shock."

2. If there is information similar to that in 5, but inadequate to impose restrictions on all columns of $H^{(-i)}$, it is still possible to limit the identified set. For example, "unemployment does not respond contemporaneously to the interest rate shock, but there are no restrictions on inflation or the interest rate." This still leaves two columns that could correspond to the interest rate.

3. The researcher may think that the response of a variable to its own named shock should be larger than the response of any other variable to that shock. Significantly, similar reasoning applies to Impulse Response Functions (IRFs), which can be directly computed and compared from the candidate H matrices and used in the labeling procedure.[7] Alternatively, some other ordering could be imposed on these objects. For example, "the monetary policy shock must have a larger impact on the interest rate residual than on any other series." Note, that this sometimes maps multiple columns to a single label, particularly if the units of the data are not comparable/normalized.

4. Imposing sign restrictions on $H^{(i)}$ or rows of H (or, again, more intuitively, IRFs) can also rule out shock labelings deemed to be unreasonable. This is much the same as the

[7] IRFs plot the dynamic causal effect of a scaled shock on a variable of interest, holding constant all other contemporaneous and future shocks.

sign restrictions identification literature pioneered by Uhlig (2005), only here it is a final step towards point identification, not the sole basis for identifying an uncountable set of candidate H matrices. For example, one could assume relatively liberally, that "the instantaneous response of inflation to the interest rate shock is positive for unemployment and negative for inflation," which might hold true for multiple identified shock series.

5. It is often appealing to impose magnitude restrictions on $H^{(i)}$ or rows of H (or, again, more intuitively, IRFs). Generally, certain scales of response are simply unrealistic. Since in practice there are often very small responses of some residuals to some shocks, any ordering that normalizes by such an element in a given column yields unrealistically large responses for other variables. For example, a researcher could assume something like "the instantaneous response of unemployment to a unit interest rate shock has magnitude less than 5 standard deviations."

6. Finally, the Sims (1980) and Blanchard & Quah (1989) logic for short-run and long-run restrictions can be extended. If no candidate ordering of columns match these assumptions exactly, the researcher can choose the columns whose implications are closest to the assumptions under some norm. For example, "after 16 quarters, the cumulative response of inflation to a unit monetary policy shock is zero."

A final approach, while not strictly an identification argument, since it references a single observed draw rather than population moments, relies on filtered volatility paths. These can be compared to the historical record to rule out elements of the identified set. For example, a high volatility of inflation shocks is expected to have occurred during the 1973 oil crisis and the Volcker period. This is essentially the "winnowing constraint" discussed above. In practice, as in the empirical application, this can be, at the very least, a convincing check on another means of shock labeling. Moreover, some researchers will likely find it to be one of the most intuitive methods of labeling the shock series. This is similar in spirit to the use of knowledge of economic events to define regimes in the Rigobon framework.

3.2 Inference on labeled columns

Importantly, inference techniques that are valid for an estimated $\hat{H}$ will also be valid for a labeled column of $\hat{H}$, denoted $\tilde{H}^{(j)}$, under standard conditions. Note that in general, the use of statistical measures to select a column of a matrix will impact the asymptotic distribution of the ultimate column estimates. However, for all methods above that map a single shock to each label, it is the case that the labeling criterion is consistent in the probability limit sense.

In this context, that means that as $T \to \infty$, the probability of selecting the correct column based on the criterion approaches unity. Pötscher (1991) establishes asymptotic distributions in a discrete model selection setting building on intuition dating back to at least Geweke & Meese (1981). The context considered here and the strong notion of consistency of the labeling criterion allow a strong form of his results to hold. Thus, if an estimator $\hat{H}$ has a known asymptotic distribution, and the labeling method is consistent, the asymptotic distribution of the labeled column will be that of the column under the distribution of $\hat{H}$. In other words, the labeling problem can be ignored asymptotically for the purpose of inference.

3.3 Transparency and reporting

TVV-ID demonstrates further value by enabling transparent discussion about the impact of economic assumptions on the estimates obtained. In particular, since the more subjective "economic" identifying assumptions are only used to label a defined set of shocks, or, equivalently, to identify which column $H^{(i)}$ pertains to a shock of interest, it is straightforward to specify what values would be identified under alternative assumptions. Thus, the notion of a result being robust to identifying assumptions is very clear. In many cases, a variety of sets of assumptions lead to precisely the same result. Showing this can make empirical work more compelling, in that a reader ascribing to any single member of that set of assumptions can be convinced by the result, even if she does not agree with the validity of all such sets of assumptions. In contrast, in much empirical work of the nature considered here, the scope for comparison of identifying assumptions has been limited. When such a juxtaposition is present, even if both sets of assumptions were valid, they would only yield quantitatively identical results in a finite sample under very specific circumstances. Further, when the subsequent results differ, it is hard to pin down exactly what aspect of the assumptions led to those differences and precisely how this influence operates. Here, on the other hand, as the impact of each assumption discriminates between a small number of discrete possibilities, it is simple to discuss.

Finally, TVV-ID ought to be attractive to researchers not wanting to take a strong stand on such economic assumptions, or wishing to leave the reader scope to decide the credibility of the identification. TVV-ID provides the option of reporting the values of $H^{(i)}$ under multiple, possibly all, column orderings. Reporting this finite, identified set is not usually an option, but here could provide a check on the integrity of the results and illuminate the degree to which the convenience of the findings may influence assumptions made.[8]

[8]Reporting all possible labelings is of course possible in any identification via heteroskedasticity approach, or others where identification is up to column order. However, what sets TVV-ID apart is that extra economic assumptions (besides the generic economic content of Assumptions A, B, & D) have not been enforced prior

4 Estimation

Having established sufficient equations to identify the structural shocks, there are many options available to apply these results to estimate the parameters of interest. In this section, I outline likelihood-based inference via Markov Chain Monte Carlo (MCMC) and describe in more detail what I call hybrid GARCH, deriving asymptotic properties. Further options exist, including GMM, variants of the GARCH-based inference of Sentana & Fiorentini and methods inspired by the infill asymptotic framework. These additional methods are described in Appendix B.2. While stationarity is not needed for identification, many estimation procedures will require the assumption for desirable asymptotic properties to hold.

4.1 Quasi-likelihood inference based on time-varying volatility

A quasi likelihood approach is appealing because it provides a natural way to incorporate the identifying information of multiple autocovariances. The drawback of any likelihood-based approach is the necessity of specifying a law of motion for the structural variance; to some extent this is a return to parametric assumptions this paper set out to avoid. However, thanks to the general identification arguments offered above, a researcher can specify any functional form for time-varying volatility provided it implies an autocovariance. It is also possible to fit multiple forms to examine how robust results are to such assumptions. Each functional form implies various moments for ζ_t, exploiting the result in Theorem 2 showing that additional moments lessen the risk of weak or non-identification. There is an extensive literature discussing functional forms for time-varying volatility in the financial econometrics literature, see e.g. Shephard (1996) or Fuh (2006). A popular general form is a simple AR(1) log SV model[9]. For a dimension i, it takes the form

$$\log\left(\sigma_{it}^2\right) = \phi_i \log\left(\sigma_{it-1}^2\right) + e_{it},$$

where e_{it} and e_{jt} can have arbitrary covariance. Provided $|\phi_i| < 1$, this provides a stationary approximation to popular models viewing log-variance as a random walk. Such general forms for the likelihood require estimation via simulation methods like MCMC. The AR(1) SV model is applied throughout this paper, and ultimately constitutes the recommended implementation for TVV-ID. The reasons for this are addressed in the simulation study of Section 5. For more details on quasi-likelihood estimation in this context, see Appendix B.2.

to this labeling problem, so it is possible to leave all "economic" decisions in the hands of the reader.

[9]Henceforth, I use the terms AR(1) log SV and AR(1) SV interchangeably to refer to the same model, with the choice depending on the emphasis in context.

4.2 Hybrid GARCH

A hybrid method based on the GARCH functional form has the advantage of minimizing nuisance parameters to be estimated and not requiring intensive algorithms like MCMC, without directly strictly imposing the GARCH structure. Calibrated GARCH parameter values can be used to form a kernel and obtain a volatility path. Quasi-Maximum Likelihood (QML) estimation can then be performed for H based on the implied filtered path. The applicability of the GARCH model is discussed in e.g. Engle (2001). Consider the GARCH(1,1) functional form for $t = 2, 3, \ldots, T$:

$$\sigma_{it}^2 = \mu_i (1 - \psi - \Upsilon) + \phi\sigma_{i,t-1}^2 + \Upsilon\varepsilon_{i,t-1}^2, i = 1, 2, \ldots, n, \text{ and } \mu_i, \psi, \Upsilon \geq 0. \tag{11}$$

The GARCH(1,1) law of motion means that if $\psi + \Upsilon < 1$, then $E[\sigma_{it}^2] = \mu_i$, where the expectation is with respect to the stationary distribution. The hybrid approach deviates from standard GARCH by fixing the values of ψ and Υ via calibration (calibration details are discussed in Appendix B.5). My focus is allowing μ_i to remain a free parameter to capture the mean variance of each series, as in Pakel, Shephard, & Sheppard (2011); experimentation suggests doing so greatly improves performance.

I now establish the asymptotic properties of the hybrid GARCH estimator. Define the filtration $\mathcal{F}_{t-1} = \{\sigma_1^2, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{t-1}\}$, $\mathcal{F}_0 = \sigma_1^2$. Stacking (11) gives

$$\sigma_t^2 = \sigma_t^2 (\mu, \mathcal{F}_{t-1}), t = 1, 2, \ldots, T,$$

where $\mu = (\mu_1, \ldots \mu_n)'$. Using a QML approach, $\left(\mu', \sigma_1^{2'}\right)'$ can be estimated simultaneously with the free elements of H. I consider the working densities corresponding to

$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, H\Sigma_t (\mu, \mathcal{F}_{t-1}) H'), \; t = 1, 2, \ldots T, \tag{12}$$

recalling $\eta_t = H\varepsilon_t$ and $\Sigma_t (\mu, \mathcal{F}_{t-1}) = diag (\sigma_t^2 (\mu, \mathcal{F}_{t-1}))$. It is straightforward to maximize the joint quasi-likelihood for $t = 1, 2, \ldots, T$ with respect to $\left(\mu', \sigma_1^{2'}\right)'$ and H. In order to obtain the asymptotic properties of this procedure, I turn to the QML literature, following White (1982). For a discussion of GARCH estimation via QML, see Bollerslev & Wooldridge (1992); Normandin & Phaneuf (2004) consider the present problem, where multiple series are related by H, using maximum likelihood.

Define θ as $\left(\mu', \sigma_1^{2'}\right)'$, plus the non-diagonal elements of H, with $\theta \in \Theta$. Let the true joint distribution of η_t be G over Ω with Radon-Nikodym density $g(\eta_t \mid \mathcal{F}_{t-1}; \theta)$. Denote the model joint density, the multivariate normal density implied by (12), as $f(\eta_t \mid \mathcal{F}_{t-1}; \theta)$. Define $L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} E_G[\log f(\eta_t \mid \mathcal{F}_{t-1}; \theta)]$, where $E_G[\cdot]$ denotes the unconditional expectation

with respect to G. It is well known that maximizing $L_T(\theta)$ with respect to θ is equivalent to minimizing the Kullback-Leibler distance with respect to θ. Denote θ^* as the unique maximizer of $L_T(\theta)$. The sample counterpart of $L_T(\theta)$ is $\bar{L}_T(\eta_t;\theta) = \frac{1}{T}\sum_{t=1}^{T} \log f_t(\eta_t \mid \mathcal{F}_{t-1};\theta)$ with maximizer $\tilde{\theta}_T$.

Consistency

To establish the consistency of $\tilde{\theta}_T$ for θ^* as $T \to \infty$, I make the following assumptions.
Assumption E.

1. Θ is compact,
2. $E_G(\log g(\eta_t)) < \infty$,
3. $0 < \sigma_1^2 < \infty$,
4. $\psi \geq 0, \Upsilon \geq 0$, $\psi + \Upsilon < 1$.

E.3-4 imply that the path of σ_t^2 is strictly bounded away from zero and is finite with probability one. This means that f is measurable in η_t for all $\theta \in \Theta$ in addition to being continuous in θ for all $\eta_t \in \Omega$. Further, $E_G|f_t(\eta_t \mid \mathcal{F}_{t-1};\theta)| < \infty$ for all $t = 2, 3, \ldots, T$ since $\sup_{\eta_t|\eta_{t-1}} |f(\eta_t \mid \mathcal{F}_{t-1};\theta)| < \infty$. Together with E.2, this last fact guarantees that the Kullback-Leibler distance is well-defined. Identification of a unique $\theta^* \in \Theta$ that minimizes the Kullback-Leibler distance is established by the main theorems of this paper. This completes the necessary conditions to apply Theorem 2.2 of White (1982), which yields a strong consistency result:

$$\tilde{\theta}_T \overset{a.s.}{\to} \theta^*$$

This shows that the QML estimator is a strongly consistent estimator for the minimizer of the Kullback-Leibler distance.

Asymptotic normality

To characterize the asymptotic distribution of $\tilde{\theta}_T$, I impose further assumptions on θ^*:
Assumption F.

1. θ^* is a regular point of $D(\theta) = E_G[\nabla^2 \log f(\eta_t \mid \mathcal{F}_{t-1};\theta)]$,
2. $B(\theta) = E_G\left[\nabla \log f(\eta_t \mid \mathcal{F}_{t-1};\theta) \nabla \log f(\eta_t \mid \mathcal{F}_{t-1};\theta)'\right]$ is invertible.

The definition of f as a multivariate normal (with E.2-3) satisfies the further properties assumed in White (1982): $\nabla \log f(\eta_t \mid \mathcal{F}_{t-1}; \theta)$ is a measurable function of η_t for each $\theta \in \Theta$; continuously differentiable in θ for each $\eta_t \in \Omega$, the sample space; and $|D(\theta)|$ and $|B(\theta)|$ are integrable with respect to G for all η_t and $\theta \in \Theta$. Then, under Assumption F, White (1982) Theorem 3.2 gives

$$\sqrt{T}\left(\tilde{\theta}_T - \theta^*\right) \xrightarrow{d} \mathcal{N}\left(0, C\left(\theta^*\right)\right),$$

$$C_T\left(\tilde{\theta}_T\right) \xrightarrow{a.s.} C\left(\theta^*\right) \text{ element by element,}$$

where $C(\theta^*) = D(\theta^*)^{-1} B(\theta^*) D(\theta^*)^{-1}$. This offers asymptotic normality of the estimator at the Kullback-Leibler minimizing value θ^*. The natural sample counterparts can be used for inference. In the case that there exists a θ_0 such that $f_t(\eta_t \mid \mathcal{F}_{t-1}; \theta_0) = g_t(\eta_t \mid \mathcal{F}_{t-1})$ for all $t = 1, 2, \ldots, T$, then $C(\theta^*)$ simplifies to the Cramér-Rao lower bound, $C(\theta_0) = -D(\theta_0)^{-1}$.

The underlying standard set of parameters used to calibrate ψ and Υ may vary based on the application and especially frequency considered – whether highly volatile financial variables or slow-moving macro variables. For the purposes of the simulation study and empirical application below, I calculate such a standard set of parameters based on the monthly 120-variable macro dataset of Stock & Watson (2002) and BBE in Appendix B.5. The following simulation study shows that hybrid GARCH estimation performs well.

5 Performance of estimators

TVV-ID performs well in comparison to alternatives. First, I present a theoretical argument highlighting difficulties encumbering the Rigobon approach when regime breaks must be estimated. Second, I present simulation results across a range of estimators. I examine the performance of a variety of popular regime-estimation approaches in simulation; the results are disappointing. Then I compare the performance of TVV-ID (under multiple estimation approaches) to various applications of the Rigobon scheme and Sentana & Fiorentini's GARCH identification. I do so across several data-generating processes (DGPs), and with varying degrees of time variation in volatility. Throughout the cases studied, TVV-ID exhibits the best performance, and, in particular, the likelihood-based estimators are superior. The hybrid GARCH method (see Section 4.2) is the best method not relying on simulation-based inference.

5.1 The breakdown of conditional diagonality

When regime breaks must be estimated, estimates obtained via the Rigobon method face an important source of bias. It is standard to assume that $E\left[\varepsilon_t\varepsilon_t'\right]$ is diagonal; the Rigobon scheme requires in addition that $E\left[\varepsilon_t\varepsilon_t' \mid t \in A\right]$ is diagonal for the sub-sample $A \subset T$. There are two potential forces driving any norm of $H\varepsilon_t\varepsilon_t'H'$ to be "high-valued" or "low-valued" – in A or not in A. On the one hand, a period of high volatility increases the norm. On the other, certain values of $\varepsilon_t\varepsilon_t'$ will be more conducive to a high norm (the precise values will especially depend on H and also on σ_t^2). Thus, conditional on being in a certain data-dependent sub-sample, some draws of $\varepsilon_t\varepsilon_t'$ will be more likely than others, and this equally applies to off-diagonal elements of $\varepsilon_t\varepsilon_t'$. Thus, $E\left[\varepsilon_t\varepsilon_t' \mid t \in A\right]$ will not, in general, be diagonal. Another way to think about this is that the estimated regimes are not exogenous to η_t as they are calculated based on η_t.

A simple numerical example illustrates this fact. For this purpose, compare $H = I_2$ to the generic $H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Let the "low variance" regime be $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the high $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$; shocks are normally distributed. I take 500,000 samples from each regime and compute the trace of $H\varepsilon_t\varepsilon_t'H'$ for each observation. Sub-sample A consists of those draws whose trace is above the overall median, and B below. Table 1 computes conditional expectations. With H as the identity, the off-diagonal elements are indistinguishable from zero, but with non-zero off-diagonal elements in H, the diagonal structure clearly breaks down, as described above.

This lack of diagonality within the sub-samples biases estimates of H. Consider the 2-dimensional case. When diagonality holds, $E\left[\varepsilon_{1t}\varepsilon_{2t} \mid t \in A\right] = 0$, so

$$\sigma^2_{\eta_{11},A} = E\left[\varepsilon_{1t}^2 \mid t \in A\right] + H_{12}^2 E\left[\varepsilon_{2t}^2 \mid t \in A\right] = c_1 + H_{12}^2 c_2.$$

Without diagonality,

$$\begin{aligned}\sigma^2_{\eta_{11},A} &= E\left[\varepsilon_{1t}^2 \mid t \in A\right] + H_{12}^2 E\left[\varepsilon_{2t}^2 \mid t \in A\right] + 2H_{12}E\left[\varepsilon_{1t}\varepsilon_{2t} \mid t \in A\right] \\ &= c_1 + H_{12}^2 c_2 + H_{12}c_3,\end{aligned}$$

which includes an additional unknown, c_3. It is clear that assuming $c_3 = 0$, as the literature does, biases estimates. The problem is compounded for higher dimensions. Simply speaking, the Rigobon argument, which yields just-identification with two regimes, is now under-identified if c_3 must be determined.

This issue is most clearly illustrated when one considers the alternative sub-sample identification argument (and its estimation analog) offered by Sims (2014). If $S_A = H\Sigma_A H'$

where $\Sigma_A \equiv E\left[\Sigma_t \mid t \in A\right]$ and similarly for B, then

$$S_A S_B^{-1} = H \Sigma_A \Sigma_B^{-1} H^{-1}.$$

If $\Sigma_A \Sigma_B^{-1}$ is diagonal, then those diagonal elements are the eigenvalues of the matrix on the right hand side, and the columns of H are the corresponding right eigenvectors (uniquely so if the eigenvalues are distinct). However, if Σ_A, Σ_B, and thus $\Sigma_A \Sigma_B^{-1}$ are not diagonal, then the diagonal elements are not the eigenvalues of the matrix, and the columns of H are not the eigenvectors. Therefore, diagonality conditional on membership in a sub-sample is crucial for estimates to be valid.

The researcher then faces a trade-off in choosing regimes. As the length of the window over which the norm of $\eta_t \eta_t'$ is computed tends to infinity, provided some heteroskedasticity is present, the off-diagonal elements will converge to zero. However, as the length of the sub-samples goes to infinity, provided stationarity holds, the covariance matrix of each subsample will converge to the same value. A weak identification problem emerges – if the covariances are identical across sub-samples, the original problem of identification only up to orthogonal rotations returns. In much macroeconomic data, from an estimation point of view, it remains unlikely that sample sizes are large enough to avoid the issue of non-diagonality within the sub-samples. A potential solution is accepting the presence of off-diagonal terms and using additional regimes to identify the extra parameters outlined above. However, in the current literature, these issues are unaddressed; further work should investigate the extent to which existing results are robust to this issue.

5.2 Simulations

I begin by describing the DGPs used across simulations. My first simulation study exposes the bias of Rigobon identification based on estimated regimes. The second study compares TVV-ID to existing identification schemes, under a variety of estimation approaches.

Data generating processes I consider a range of models of heteroskedasticity prevalent in the literature. In particular, I consider a generic AR(1) SV model, a GARCH(1,1) model, and a Markov switching model, intended to mirror the regime-based heteroskedasticity of Rigobon. Parameters are calibrated based on a version of the empirical application estimated using only two factors (and the FFR). This two-factor version was chosen so a simple two-dimensional simulation study could be calibrated to a two-factor representation of macroeconomic shocks. A summary of the data generating processes is given in Table 2; calibration details are in Table 3.

The values for the Markov regimes are inferred based on the AR(1) DGP, except for the Markov transition matrix. Fitting the model resulted in almost constant switches, whereas the goal of these simulations was to assess performance in a context like that supposed by Rigobon (2003), even if that may not accurately characterize the data studied in this paper. Thus, the probabilities were altered to lengthen regimes.

The "weak" calibration parameters are chosen to scale down the innovation variances by a factor of 100 for the AR(1) and the conditional heteroskedasticity by a factor of 10 for GARCH(1,1)[10]. These values are admittedly arbitrary, but were chosen after comparing sample variance paths under different calibrations to obtain paths that exhibited modest fluctuation in the neighbourhood of the process mean. Representative sample paths are given in the Online Appendix D.1. Throughout, H is taken as

$$H = \begin{bmatrix} 1 & -1 \\ 0.55 & 1 \end{bmatrix}.$$

The H calibration is lifted from the relationship between shocks to the two macro factors in the AR(1) MCMC estimates in the simplified empirical application. Unless otherwise noted, shocks are standard-normally distributed. For the study of Rigobon regime estimation, I take $T = 2000$, substantially longer than in the empirical application, to distinguish between true bias and fundamentally small-sample phenomena. The basic sample length for the main study, on the other hand, is shortened relative to the empirical application, to $T = 200$, because the sample studied there, with around 500 monthly observations, is quite long in the context of macroeconomics; the shorter sample puts the methods through a sterner test in terms of strength of identification. Simulations are conducted with 10,000 replications. Labeling of the columns of H proceeds via an infeasible method of comparing the L_2 norm of the resulting H matrices to the true matrix. This is infeasible because, in practice, the true H is unknown. Nevertheless, the procedure is still imperfect, as the full distributions of estimates make clear.

Study 1: Estimating regimes As a first study, I apply a range of norms, window lengths, and threshold rules motivated by those used in the empirical literature to assess the performance of the Rigobon methodology. For norms, I consider both the trace of $\eta_t \eta_t'$, and the diagonal element expected to be most impacted by heteroskedasticity, as in Rigobon & Sack (2003). For windows, given that the calibration is to monthly data, I consider single-period, 7-period, and 13-period symmetric windows. For thresholds, I consider both the median, which maintains precision in the estimation of both subsample covariances, and

[10]Note that for both AR(1) log SV and GARCH, this also implies a change in $E\left[\sigma_t^2\right]$.

one standard deviation above the mean, as in Rigobon & Sack (2003). I analyze both the Markov switching DGP, as it is the leading case for the Rigobon scheme, and the AR(1) SV DGP, as representative of a time-varying volatility model. Estimation proceeds using the Sims methodology discussed in Section 5.1.

For both DGPs, the majority of tuning parameters result in moderate to severe bias. For the Markov-switching DGP, the identification scheme should perform relatively well, since there truly are windows of high and low variance. Results for "oracle" estimation, where the true break dates are known, are accurate; the deviations from true values are due to the column labeling problem of the H matrix. Histograms for the oracle are available in the Online Appendix, Figure 22. Table 5 shows that performance is better for longer windows, and for the trace as opposed to a single diagonal element. For some tuning parameters, the true value is not even contained within the $[0.025, 0.975]$ quantiles of the draws. That longer windows better identify the true parameters accords with theory; here, the off-diagonal bias will be minimized. Overall, results are quite sensitive to window-length and the norm used, with the threshold having a relatively small effect.

Surprisingly, performance for the SV DGP is slightly better. Here, longer windows yield estimates quite close to the true values. The table for SV and histograms depicting estimates for both DGPs are available in the Online Appendix in Table 13 and Figures 23 & 24. In additional simulations, as the variance of SV increases, the bias shrinks, particularly for long windows. However, the variance must increase by several orders of magnitude before the parameters of H can, in general, be deemed "well-estimated". Regardless, for empirical calibrations, estimates based on sub-sample identification are both susceptible to bias and sensitive to tuning parameters. Even if the researcher has a strong belief that the underlying heteroskedasticity is dramatic, such methods must be applied cautiously.

Study 2: Comparison of estimators Three identification schemes are compared in this simulation study: the Rigobon approach, Sentana & Fiorentini's identification using instantaneous covariance estimates, and TVV-ID. Table 4 summarizes all of the estimators considered. The first implementation of Rigobon uses sub-samples defined based on a 13-period rolling window of $trace\,(\eta_t \eta_t')$, with the high volatility regime being defined as one standard deviation above the mean. Second, it is implemented with the two sub-samples corresponding to simply the first and second halves of the sample. The standard implementation of Sentana & Fiorentini's (2001) scheme uses maximum likelihood on GARCH(1,1) variance processes. I also use an Epanechnikov kernel smoother to estimate instantaneous covariances with bandwidth equal to 12 and then apply minimum distance estimation for H (see Appendix B.2.4 for additional details). A middle-ground between Rigobon and Sentana

& Fiorentini splits the data arbitrarily into 8 consecutive blocks of the same length and applies minimum distance estimation for H.

TVV-ID is implemented in four ways. First, an approximation to quasi-likelihood is estimated via Hamiltonian MCMC with flat priors and an AR(1) log SV model, allowing for correlated innovations across dimensions. The same model is also estimated in a Bayesian framework with weakly informative generic priors ($\mathcal{N}(0, I)$ on H, $unif(-1, 1)$ on autoregressive parameters, Lewandowski-Kurowicka-Jo prior with shape parameter 1 on the innovation correlation matrix, positive Cauchy with location parameter zero and shape parameter 2 on the variances).[11] GMM is applied using the mean and first autocovariance of $\eta_t \eta_t'$; the standard two-stage procedure is used for weighting. Finally, I estimate the hybrid approach where the GARCH parameters are calibrated as $\phi = 0.6378, \Upsilon = 0.2101$ (estimated based on 120 macro time series, see Appendix B.5), with the remaining estimation, including each process's mean, via QML.

In all, ten DGPs are considered: a version of AR(1) SV where only one dimension exhibits SV, empirically calibrated and "weak" versions of AR(1) SV and the GARCH(1,1), the Markov switching model, varying sample lengths for the empirically-calibrated AR(1) SV, and t_1 shocks for the empirically-calibrated AR(1) SV. The single-dimension SV model helps abstract from the shock labeling problem as here the estimated series are pre-labeled by assumption. This DGP also admits the closed-form ratio estimators discussed in Section 2.1 for both sub-sample approaches and the autocovariance. Since this paper recommends a parametric approach based on an AR(1) SV volatility model, while being critical of previous work assuming GARCH, a key goal is to compare the performance of those estimators under misspecification. As before, labeling is conducted using the infeasible method based on the L_2 norm. This procedure allows the results to focus on the challenge of estimating H, as opposed to that of labeling the shocks thereafter.

The median values obtained from each estimator as well as coverage rates of the associated standard errors are reported in Tables 6 & 7. Histograms for the estimation of H_{12} in the basic AR(1), $T = 200$ DGP and the Markov switching DGP are presented in Figures 2 & 3. Histograms for all estimates and DGPs are available in the Online Appendix, Figures 25-34. The over-arching results, visible from the included Tables and the histograms, follow. The likelihood estimators based on the AR(1) SV process perform best across the various DGPs used to generate the data. While small biases can occur, the central tendency remains quite reliable except for when confronted with a weakly-identified GARCH process, at which

[11] While it might seem plausible that these priors are too informative (particularly on H) they accord well with most economists' strong intuition for the size of the elements of H with a unit diagonal normalization and correct labeling. Moreover, the priors have little impact on the results in practice, which remain very similar to those obtained with flat priors.

point all other estimators struggle too. It is also more efficient than other estimators. The size distortions are are generally in the 5-10 p.p. range when it is well-specified, naturally growing if it is misspecified. The hybrid GARCH estimator is the next best, and the best to not require MCMC techniques for implementation. It performs well for all but the weakly-identified DGPs, the Markov Switching process, and non-Gaussian innovations. Otherwise, it is only slightly less efficient than the likelihood methods. Size distortions are generally in the 5-15 p.p. range except for the weakly-identified DGPs.

The performance of the other estimators is notably weaker. The block-based estimator performs similarly to the hybrid, except that it appears slightly less efficient and is biased for the GARCH and single-dimension SV DGPs. Size distortions are prohibitive except for large T. The performance of GMM is generally similar to that of the hybrid, but less efficient, and it breaks down more severely when faced by weak identification. Size distortions are generally substantial, except for very large T. The superior performance of the likelihood estimators compared to GMM is due to the difficulty of obtaining a global minimum with any certainty due to the geometry of the objective function and high-dimensional parameter space, even when starting values are the true values (as was the case in this study). The smoothing applied by the likelihood function in the quasi-likelihood AR(1) SV and Bayes implementations greatly mitigates this problem. The GARCH estimator is comparable to the hybrid and GMM except for the fact that for most DGPs, it estimates substantial excess mass for H_{21} and H_{12} around zero. This pathological feature appears to result from the likelihood maximization finding local minima at the edges (i.e. near-explosive region) of the GARCH parameter space. For near-explosive dynamics, H is forced towards zero to match the sample properties. This problem is not faced by the hybrid GARCH method, since the GARCH parameters are fixed via calibration. The two Rigobon estimators, a 13-period rolling window and a simple half-sample split, perform comparably across DGPs. They are less efficient than other estimators, and face substantial bias under GARCH, Markov switching, and any weak identification. The ad hoc block bootstrap adopted for standard errors is severely undersized. The kernel estimator performs erratically, which makes it hard to draw any general conclusions.

6 Empirical application: Bernanke, Boivin, & Eliasz (2005)

TVV-ID can help assess the reliability of important results in the monetary policy literature. In particular, much research imposes a recursive form on the H matrix, which TVV-ID does not require. Here, I consider the influential Boivin, Bernanke, & Eliasz (2005) paper, which develops the FAVAR methodology to make an important methodological contribution as well

as substantive statements about monetary non-neutrality. I begin by offering background on BBE and the FAVAR approach. I then discuss results obtained using TVV-ID and compare them to those of BBE and the broader literature. I finish by offering three extensions to the FAVAR framework of BBE. I apply TVV-ID to a longer alternative dataset, consider two-dimensional monetary policy to study "guidance" effects, and perform sub-sample tests to appraise the stability of H from 1960-2000.

6.1 Background

BBE's FAVAR methodology addresses several econometric concerns present in SVARs. A key assumption of the SVAR methodology is that the reduced-form residuals, η_t, span the structural shocks, ε_t. If not, the system of equations determining H suffers from omitted variables problems. To this extent, including additional variables in a VAR is seen as important. Unfortunately, those macro time series are frequently very noisy. Measurement error abounds, and one need only look at the frequency of revision of time series to recognize that these may not be the best measures for solving the omitted variables problem. The FAVAR, instead of including a few series in the VAR, estimates the most important common factors of a potentially large set of variables and includes these in the VAR. If enough underlying series are used, and measurement errors are not highly correlated, these estimated factors should not suffer from the same errors, and should also incorporate information from a wider set of variables than those often included in a small-scale VAR.

BBE apply this methodology to study the impact of Federal Funds rate (FFR) shocks from 1960-2001. They use a dataset of 120 monthly macro variables from Stock & Watson (2002). For a detailed discussion of their procedures, the reader should consult the original paper. For most specifications, they include only the FFR as an observed series in the VAR. They then estimate the common factors on the remaining macro series. Common factors are also estimated on only "slow moving" time series. In an effort to remove contemporaneous dependence of the factors on the FFR, they regress the full set of factors on these slow moving factors and the FFR and subtract off the portion predicted by the FFR. Note that while this is important to add basic credibility to the recursive (Cholesky) ordering they impose later, it is not necessary under alternative identification schemes, but I maintain it for consistency.[12] BBE then estimate a VAR on the orthogonalized factors and FFR and compute structural impulse response functions (IRFs) using a Cholesky decomposition, ordering the FFR last. Responses for any other variable in their dataset can be constructed by regressing its path on that of the factors and FFR, and using these coefficients to predict its impulse response

[12] I abandon this procedure when I perform the extension to the McCracken & Ng.

based on those of the factors and FFR. Standard errors proceed via bootstrap, which must account for the generated regressors problem arising from estimated factors.[13]

I follow BBE, up to the decomposition of shocks, at which point I apply TVV-ID. While they use a Cholesky decomposition to estimate H and construct structural IRFs, I apply TVV-ID, fitting an AR(1) SV model via MCMC. I label the monetary policy shock using a zeroth-order FEVD: I assume that it will be the best predictor of reduced-form residuals in the FFR. In this context, this amounts to saying that while there is likely a non-zero feedback from unexpected contemporary movements in macro variables to innovations to the FFR, these effects are not responsible for the majority of innovations to the FFR. Admittedly, this assumption is subjective, but further evidence supports it, as discussed below.[14] The assumption is also related to Ramey's (2016) finding that large movements in the non-systematic component of monetary policy are increasingly rare. She maintains that this is due to the fact that there are few true monetary policy shocks in recent times, with most changes labeled as such instead being effects of "superior information on the part of the Fed, foresight by agents, and noise". It is hard to construe any such movements as fundamental macroeconomic shocks. It is worth noting that under TVV-ID, it is not necessary that there actually be large shocks to the series of interest, merely that the volatility of some shocks exhibit movement; as such, Ramey's concern for identification in recent decades can be somewhat assuaged.

My inference methods also differ from BBE. For the IRFs, my standard errors are computed via a hybrid method exploiting the block diagonal structure of the variance of the component of the structural IRF estimates, incorporating the block bootstrap of Gospodinov & Ng (2013). This involves resampling from the 120 time series, estimating new factors for each draw, and re-estimating the AR coefficients and the coefficients relating external regressors to the variables included in the VAR. For each draw, the reduced-form IRFs for any variable of interest can be constructed at each horizon. The covariance of these draws of the reduced-form IRFs can then be computed. A covariance matrix for the MCMC estimates of H is computed using Müller's (2013) sandwich adjustment to account for the possibility of misspecification. Using the well-known block-diagonal structure of the covariance matrix for reduced-form coefficients and structural parameters (eg. Lütkepohl, 2006), the delta method can be applied to obtain the variance of the structural IRFs. For more details, see Appendix B.4. For ease of comparison, I replicate BBE's baseline results with standard errors calculated using this procedure, depicted in Figure 38 of the Online Appendix. It

[13]While Bai & Ng (2006) offer conditions under which the generated regressors problem can be ignored for the purpose of standard errors for factor models, the number of series is too small in this application relative to the length of the time series for such results to hold.

[14]See Section 3.1 for a detailed discussion of alternative shock labeling schemes.

is worth noting that these standard errors, are, on the whole, smaller than those of BBE. BBE provide few details of their bootstrap method, which appears based on an approach originally developed to address concerns of bias in OLS estimates in small samples. The fact that some of the resulting bands are very narrow for BBE's Cholesky approach is natural – it is rare to estimate AR coefficients with large errors in a sample of this length.

6.2 Results

TVV-ID produces compelling results in support of monetary non-neutrality. As an initial step, the results of BBE were successfully replicated, using data available from RePEc.[15] This ensures that any differences observed are due to the identification of H. Following the results of my simulation study, my baseline estimates come from the AR(1) log SV quasi-likelihood approach; additional results using the hybrid estimator are reported in Online Appendix D.3, and are much the same, as are unreported results using other estimators.

It is necessary to label the monetary policy shock. Table 8 reports the candidate values for $H^{(FFR)}$. I consider the zeroth order FEVD via the marginal R^2.[16] This selects the fourth column reported (in purple). Note that while the scale of the factors themselves and thus these coefficients are not immediately interpretable, the factors are standardized and estimated on standardized data. Thus, responses of the orders of magnitude of the other columns are largely incomprehensible in relation to macroeconomic theory. The marginal R^2 corresponding to the fourth series is greater than 0.99 (note that the overall R^2 is mechanically unity), meaning that structural shocks besides the monetary policy shock have virtually no impact on the FFR. This also implies that in this setting, the reduced-form innovations to the FFR are basically equal to the latent monetary policy shocks. This is reminiscent of a Cholesky ordering with the FFR first, instead of last, although that structure does not extend to the remaining rows of H. This finding is discussed in more detail below.

The implied paths for the variance, presented in Figure 4, are also informative. They are demeaned and normalized by standard deviation to render them comparable. Note that these plot the shock variance, not the shocks themselves. If a macroeconomist were asked, absent any data, to make a single prediction about the variance of monetary shocks during 1960-2001, it would surely be that it experienced a precipitous rise at the time of the Volcker disinflation (and subsequently fell through the Great Moderation). Accordingly, the labeling selected by the FEVD chooses the one volatility path that matches these predictions. This

[15] Many thanks to Tom Doan for making this data available on RePEc. https://ideas.repec.org/c/boc/bocode/rtz00012.html

[16] The marginal R^2 for ε_{it} is defined as the $R^2 - R^2_{-i}$, where R^2 is based on all $n$shocks, and R^2_{-i} is based on all shocks but i. Since $R^2 = 1$ mechanically in this model, the marginal R^2 simplifies to $1 - R^2_{-i}$.

offers a satisfying sanity check on the estimation results, and suggests that the FEVD does a good job labeling the shock series.

While Figure 4 shows a single period of high volatility for the monetary policy shock, that does not mean there was only one notable monetary shock. This is illustrated by Figure 5, which plots the monetary policy shock series against the FFR. While the largest shocks are during the high-volatility period, there are sizable shocks in the 1970s and later 1980s, before the pronounced moderation in the 1990s. Figure 6 shows that the path of the shocks is qualitatively similar to that constructed by Romer & Romer (2004), where both series are standardized; the correlation is 0.43. Volatility is actually higher in the Romer shocks, but that is likely due to the fact that the information set to which they are orthogonal (Greenbook forecasts) is smaller than the underlying set here.

The presence of a prominent period of high volatility in the monetary policy shock seems to suggest using Rigobon's approach, but that intuition has limitations. Recall that under TVV-ID, identification of the monetary policy shock is not based on the volatility of that shock alone, but rather on the time-varying volatility of the entire system. Thus, it also exploits the more frequent fluctuations in the other series, which would be largely ignored by a Rigobon approach centered on the Volcker period. Indeed, applying Rigobon around the Volcker period (defined as the dates that the volatility path first exceeds and then falls below its mean in the relevant date range) delivers qualitatively similar conclusions for the monetary policy shock column of H, although the magnitudes differ markedly. However, the results for the other columns are not as consistent; this is in keeping with the fact that for other series, the variances are likely comparable between the Volcker and non-Volcker periods, leading to a weak identification problem, as discussed in Lewis (2017). The issue is further compounded by the fact that the Volcker regime is estimated on only 31 observations. Graphically, Figure 7 superimposes those high variance regimes chosen by a standard Rigobon rolling window on the variance paths estimated by TVV-ID. There are no high-variance episodes in the second half of the sample, and the charge that regimes greatly smooth away potential identifying information is supported.

Tests using these results rejects BBE's Cholesky assumptions on the monetary policy column of H. The tests are presented in Table 9. While the size of the coefficients themselves is not very meaningful as they pertain to arbitrarily rotated factors, it is notable that none appears near-zero. The joint Wald test rejects Cholesky structure at all conventional levels. Turning to IRFs allows a more meaningful interpretation of these findings.

Comparing the core IRFs (FFR, Industrial Production (IP), Consumer Price Index (CPI)) to the BBE baseline shows similar qualitative results. Figure 8 is comparable to Figure I in BBE. It plots responses to a 25 basis point monetary policy shock, where the responses are

measured in standard deviations of the response variable (e.g. first differences of $\log(IP)$ and $\log(CPI)$). It compares their baseline 3-factor Cholesky IRFs to 3-factor TVV-ID. The TVV-ID results are different, but exhibit much the same behaviour. An important exception is the price puzzle, which is discussed in detail below. A surprising result is an initial increase in IP following a positive Fed Funds shock, before the economy enters recession. TVV-ID IRFs can also provide a check on the column labeling approach. Figure 9 depicts the IRFs corresponding to the alternative labelings; they can be ruled out by any plausible magnitude restrictions. This corroborates the labeling arguments made for H above.

TVV-ID presents evidence of monetary non-neutrality for additional macro variables. I replicate Figure II in BBE, plotting a host of macro variables' response to a 25 basis point monetary policy shock. The results for my baseline are displayed in Figure 10 with 95% confidence intervals. On the whole, the results are very similar to those reported in BBE. Importantly, as in BBE's original work, I obtain strong evidence in favor of monetary non-neutrality at various horizons. This is seen for all macro-financial variables, consumption variables, industrial production, inflation, employment, manufacturing variables, housing starts, and consumer expectations. If anything, these results are even more conclusive than those of BBE, due to the tighter error bands computed based on the Gospodinov & Ng (2013) procedure.

Ramey (2016) observes puzzling positive movements in IP and CPI following a contractionary monetary shock when she relaxes Cholesky assumptions. She analyzes both the Romer & Romer (2004) narrative monetary shock setting and an external instruments model. For the Romers' approach, she finds a pronounced price puzzle, lasting up to two years, and up to six months of stimulus to industrial production; the results are even more long-lived with external instruments. My results corroborate Ramey's suggestion that the Cholesky assumption has played a large role in previous work's claims to resolve the price puzzle; the TVV-ID results exhibit a more substantial price puzzle than the BBE baseline. Without those zero restrictions, the price puzzle remains quite pronounced, though less so than in a standard 3-variable monetary policy VAR (Cholesky ordering of IP, CPI, FFR).

To help understand how Cholesky structure may mitigate these puzzles, Figure 11 displays the 25 highest magnitude initial responses of macro variables (in standard deviations) to a 25 basis point monetary policy shock. Unsurprisingly, most of the largest responses are in interest rates and spreads. Nevertheless, several macro variables – of the sort other work assumes to have no initial response – feature in the list. Many producer-side variables respond immediately, including help-wanted ads, deliveries, loans, and inventories. In addition, unemployment duration falls. There are also responses in some price indices and housing measures. However, all of these initial responses in non-financial variables are in

the opposite direction of what theory suggests. These initial responses are all less familiar forms of Ramey's puzzles. Yet, with a Cholesky structure, their responses via factors are all assumed to be zero.

In terms of scale, while the core TVV-ID responses display much the same puzzles Ramey identifies, they do so to a smaller extent, especially for IP. The price puzzle's length is similar to Ramey's results based on the Romer & Romer shocks. It is less pronounced than that of Barakchian & Crowe (2013), comparable to Coibion (2012), and only modestly longer than that obtain in the Christiano, Eichenbaum, & Evans (1999) benchmark, even though all three studies exploit Cholesky decomposition.[17]

The fact that FFR innovations seem to have no contemporaneous response to "macro" shocks merits comment. It is important to differentiate these within-month responses from responses that are present systematically in the data, as the FOMC observes macroeconomic conditions and responds to them with a natural policy lag. This also aligns with the sentiment that monetary policy decisions are frequently discussed at the preceding FOMC meeting. Additional analysis suggests that the macro structural shocks are not well-predicted by any of the data in the Fed's information set that is likely available with any accuracy within the month (financial indicators and other series computed with high frequency). This is unsurprising, as the factors in the VAR are designed to capture "slow moving" variation. Figure 12 plots the forecast error variance decomposition of FFR innovations due to monetary policy shocks and macro shocks (combined) over various horizons. It is clear that while at a short horizon, virtually all of the variation is explained by monetary policy shocks, that trend is dramatically reversed over time, with macro conditions determining much of the movement in the FFR in the long run. In light of more recent research discussing the presence of multiple dimensions of monetary policy, it is worth considering whether some form of policy (like guidance) other than FFR changes responds more contemporaneously. This is explored below.

Figure 12 plots the forecast error variance decomposition of innovations to the FFR, IP, and CPI over various horizons. This shows that the monetary policy shocks identified by TVV-ID do in fact matter. Not only do they explain movements in the FFR, as noted above, they account for a substantial share of the variation in IP, 40% three years out. The same is not true for CPI, for which the impact peaks at under 10%, after five to six months.

From a theoretical perspective, producer-side variables exhibit initial movement, notably new orders and housing starts. On the other hand, responses in consumption are much

[17]The reason the puzzles are not further reduced in these papers, in light of BBE's results, is likely due to the fact that instead of using macro factors, they select a few judicious controls. The ability of Cholesky to eradicate the price puzzle appears linked with the usage of factors.

more gradual, as expected in lifecycle models of consumption. I also find evidence of an "employment puzzle", where the initial response of employment and unemployment are in the expansionary direction, which has not been discussed in the literature. In the face of such fluctuations, wages remain sticky, as theory suggests. Significantly, the size of my core responses is quite similar to those obtained in Christiano, Eichenbaum, & Evan's (2005) seminal paper. The IP responses align best with the authors' variants with no habit formation and first-order adjustment costs. The CPI responses align best with first-order adjustment costs. Compared to alternative empirical models, TVV-ID rejects mechanical restrictions on the speed with which variables can respond to shocks. It is intuitive that theoretical models with fewer forms of inertia, and thus a better ability to also accommodate contemporaneous movements, should better match the results of TVV-ID.

6.3 Extensions

I now extend the BBE framework to further characterize the impact of monetary policy post-1960. I begin by considering a longer sample, investigate the impact of "guidance" shocks, and test the stability of H.

Updated sample

The sample considered by BBE ends in 2000. Extending the sample through the present may be problematic due to the 2008 crisis, but augmenting series through the end of 2007 can help to further establish the robustness of these results. Instead of the original BBE dataset, I consider the FRED-MD dataset of McCracken & Ng. I limit it to 123 series from 1/1960 to 12/2007 due to missing data. Note that McCracken & Ng recommend more comprehensive data transformations than BBE.[18] This runs the potential risk of obscuring long-run relationships between variables (see e.g. Banerjee, Marellino & Masten, 2016) in pursuit of stationarity, which may not be necessary (see e.g. Sims, Stock & Watson, 1990). Thus, besides an extension of the data, this provides a robustness test in terms of data transformations. As the goal is to apply TVV-ID, I do not use the slow-moving factors approach BBE adopt to support their recursive identification.

The first column of estimates in Table 10 reports the estimates of the selected column of H, with a marginal R^2 of 0.99. Again, the Cholesky structure is decisively rejected. Note that this is a test of a weaker hypothesis than in the main application, as I allow the factors to incorporate information on all (possibly fast-moving) series. Figure 13 displays

[18] These leave only 40 of the series non-stationary according to KPSS tests, whereas in BBE 80 are non-stationary.

the core IRFs for a 25 basis point monetary policy shock with both one standard deviation and 95% confidence intervals (all in red, with dashed and dotted intervals). Overall, the point estimates closely replicate those from the BBE data. For comparison, two alternatives are plotted. First, in blue, the BBE approach using slow-moving factors and recursive identification maps very closely to the TVV-ID responses. The results of a recursive ordering without using just slow-moving factors in the FAVAR are plotted in yellow. The confidence intervals are wider for all TVV-ID results. For the FFR and CPI, this can be related to the fact that they enter the regression in first and second differences respectively (an extra difference over the BBE data), and the cumulation for responses results in accumulated standard errors. In general though, using highly transformed data is likely to result in noisier estimates for the system as a whole.

Two-dimensional monetary policy

Since BBE (2005), a substantial body of work has considered the possibility of monetary policy having multiple dimensions. In particular, even before the forward guidance episodes of the 2008 recession, there has been speculation that the Fed's statements about future policy could have meaningful effects, besides those of contemporaneous policy changes. Starting with Gürkaynak, Sack, & Swanson (2005), researchers have incorporated (either explicitly or implicitly) an additional "policy" factor, which should account not only for changes in the current FFR but also changes to expectations about its future path, through the term structure. I operationalize this by adding the 5-year treasury yield to the FAVAR as a second observed series. The identification scheme is the same as before, with columns labeled using the marginal R^2. Note that there is potential to confound "guidance" shocks with other shocks that explain movements of the 5-year treasury yield, orthogonal to actual FFR changes. A leading example is the term premium, which is incorporated into longer-term rates. However, the term premium often incorporates aspects like inflation risk (Wright, 2011), suggesting that term-premium shocks are likely to push in the opposite direction as a guidance shock. Thus, if inflation is seen to fall in response to what is labeled as a guidance shock, it is unlikely that the shocks are contaminated by such effects.

Column 2 of Table 10 reports the estimates of the interest rate policy column of H and column 3 the "guidance" column. The columns are selected with marginal R^2 of 0.99 and 0.88. The Cholesky structure is decisively rejected for each. Interestingly, as with the FFR in the main study, there is virtually no contemporaneous response of the 5-year Treasury innovations to macro shocks; however, there naturally is a non-zero response to FFR shocks. Figure 14 plots the variance paths for the structural shocks. Note that, as before, both monetary policy shock variances follow a distinct Volcker/Great Moderation pattern. While

the guidance path largely traces that of the FFR shocks, it does show deviations, particularly exhibiting higher variance at some points during the Great Moderation. Figure 15 displays the IRFs for a 25 basis point FFR shock and a 25 basis point guidance shock to the FFR, 5-year Treasury yield, IP, and CPI, with both one standard deviation and 95% confidence intervals. The responses to the FFR shocks are very similar to the previous baseline results. However, some results lose statistical significance (in particular IP at various horizons) due to lower precision in estimating this higher-dimensional model. The point estimates of the responses to the guidance shock are as theory predicts. The response of IP is very similar to that for the FFR shock, a clear recessionary effect. The deflationary impact is smaller, but the price puzzle is diminished. However, the estimates are very noisy, and there is no statistically significant effect on these macroeconomic variables.

Panel 1 of Figure 16 plots FEVDs for the FFR and 5-year Treasury due to their "own" shock versus the cumulated three macro shocks. The general pattern is the same as in the previous analysis, with the innovations to both the FFR and the 5-year yield being mainly driven by their own shocks at short horizons, and, with time, the forecast errors becoming dominated by realized macroeconomic shocks. Interestingly, the FEVD with respect to macro shocks for the bond is shifted to the right; one interpretation of this is that since long yields mainly move with information about the future, the variance explained by realized macro shocks is lessened. Finally, panel 2 plots FEVDs for IP and CPI to both shocks. While FFR shocks play an important role in IP (and less so in CPI), guidance shocks explain much less variation overall. The role of FFR shocks is smaller than in the baseline analysis, suggesting that the role of guidance may have been included in the previous estimates.

These findings contribute to the inconclusive literature on the impact of forward guidance, which juxtaposes theoretical results showing strong effects on the macroeconomy (e.g. Eggertsson & Woodford, 2003, McKay, Nakamura, & Steinsson, 2016) with mixed empirical findings of the effects on asset prices. (e.g. Gertler & Karadi, 2015, Moessner & Nelson, 2008, Swanson, 2016). In particular, Gertler & Karadi (2015) find that guidance is an important part of policy transmission, but Swanson (2016) argues that it had little to no effect on relevant asset prices during the crisis. However, these papers are typical of virtually all prior work on policy guidance in that they consider only its impact on asset prices, usually due to the structure of the model estimated. By exploiting the FAVAR approach, I offer some of the first indications of impacts on a wider range of macroeconomic variables. Guidance shocks are shown to have the expected effects, but those effects are neither statistically significant, nor do they explain a substantial amount of macroeconomic variation.

Stability of H

A crucial assumption of TVV-ID, discussed in detail in Section 2, is the stability of H. A natural point for a structural break in macroeconomic relationships in the 1960-2000 period is the end of the Volcker period, before the Great Moderation. As such, I divide the BBE sample into two subsamples: 1/1960-12/1983, and 1/1984-7/2001. Estimation takes place entirely separately, based on factors estimated across the full sample. This means that the reduced-form VAR coefficients are different in the subsamples, and the underlying volatility process is allowed to change as well. Columns 4 and 5 of Table 10 display the corresponding estimates for the monetary policy column of H for each subsample. To test for equality, four tests are considered, with the results reported in Table 11. I consider two covariance matrix estimates – one an estimate of the Fisher Information, valid if the model is correctly specified – and the more conservative quasi-likelihood counterpart of Müller (2013). I test the equality of the entire H matrix (a 12-restriction test), which is required by the identification argument, and the equality of the monetary policy column, which is of more applied interest.[19] For both classes of covariances and both hypotheses, the $p-$values are all around 0.8 – there is no evidence to reject the null hypotheses of equality across periods. Statistically, H appears to be stable, vindicating this identifying assumption.

A driving force behind this finding is likely that the latter sample appears to suffer from weak identification. Two of the estimated variance paths show very little variation. The performance of the estimator suggests that the likelihood is relatively flat in the region of the true parameters. Since estimation proceeds via MCMC, the resulting sampling variation inflates the standard errors at least to some degree, unlike in a method of moments setting, where explicit efforts must be made to ensure inference is robust. As a result, the weakness of identification may be driving the inability to reject the null hypotheses of stability. This relates to Ramey's (2016) argument that recent decades have seen minimal identifying variation for monetary policy shocks.

I now consider whether there are economically important changes in H. Figure 17 plots the core IRFs for each subsample, along with the full-sample IRF. Economically, the FFR and IP responses are very similar. The CPI paths for the early period and the full sample are also similar. However, the path of CPI for the latter sample is markedly different (with a large price puzzle). This is actually due to shifts in the reduced-form relationships between the data as opposed to structural changes in H; this is confirmed by constructing alternative IRFs

[19] To facilitate the tests, I assume independence across the two samples, which is restrictive due to temporal interdependence. However, note that the estimators are most likely to be positively correlated across samples (particularly considering the small observational overlap at the break), which can be expected to bias the test towards rejecting the null hypotheses considered.

using the split-sample H estimates and the full-sample AR coefficients (not reported). Thus, economically, there also does not appear to have been much of a shift in the transmission of monetary policy (at a structural level) across the sample.

However, the relative importance of the shocks has changed over the sample. The first panel of Figure 18 displays how the FFR has responded to monetary policy shocks versus macroeconomic factors for both subsamples. The results are qualitatively the same as the full-sample analysis, with most unpredictable movements in the FFR explained by monetary shocks at short horizons, with macroeconomic shocks coming to dominate as the horizon increases. The second panel plots the FEVDs for the FFR, IP, and CPI in response to monetary shocks for each subsample. The results for the pre-Volcker period are comparable to those for the full-sample, with a significant portion of variation in IP being explained by monetary policy shocks, and only a small portion of the variation in CPI. In the latter half, monetary shocks explain very little unpredictable movement in either series. These results can be compared to those Boivin, Kiley, & Mishkin (2010), who also take a split sample approach. They omit the Volcker period, estimating a FAVAR on the prior and following samples. They find that in general, responses are muted in the latter half of the sample. While it is hard to speak to these findings given the ambiguous responses in the latter sample here, the FEVDs bear out their results. It is clear that, as noted by Ramey (2016), among others, the role of monetary policy diminished in the latter part of the 20th century.

7 Conclusion

This paper develops a general framework under which latent variables models can be identified via time-varying volatility. The previous literature offers identification arguments based only on parametric assumptions on the variance process. In particular, I show that when regime dates are estimated, Rigobon's (2003) subsample methodology can suffer from substantial bias. In this context, I offer an identification argument that makes minimal assumptions on the variances as a stochastic process. This extends results like those in Sentana & Fiorentini (2001) by freeing the researcher from needing to assume a particular functional form (or, indeed, any functional form). Then, economic information usually used to identify the model need only be used to label the shocks. A variety of estimation methods are proposed. Simulation evidence shows that quasi-likelihood methods based on an auto-regressive log-variance process work well even when the true process has a different form.

An empirical application to Bernanke, Boivin, & Eliasz (2005) rejects the Cholesky assumptions made on the structural matrix in the original paper, but confirms and indeed strengthens the general conclusions of monetary non-neutrality. However, the price puzzle

remains when Cholesky assumptions are supplanted by TVV-ID, in keeping with Ramey's (2015) findings. While the conclusions do not differ substantially from existing work, they are obtained under weaker assumptions, and as such can be considered more robust. These results hold in an updated sample. The effects of FFR shocks are unchanged after allowing for a guidance dimension to policy, but the impacts of guidance shocks are neither statistically nor economically significant. Further, the structural matrix, H, appears to be stable over the sample, vindicating the identification assumption that H is fixed. The weak assumptions required for identification mean TVV-ID can be used in a wide array of contexts where volatilities of unobserved series are thought to vary. Although economic assumptions are used to interpret results, their impact is highly transparent. Together, these attributes suggest this method can mitigate claims that empirical macroeconomic results hinge on an authors' particular choice of assumptions.

References

Ackerberg, D. A., J. Geweke, and J. Hahn (2009): "Comments on "Convergence Properties of the Likelihood of Computed Dynamic Models"," *Econometrica*, 77, 2009–2017.

Andrews, D. W. K. (1983): "First Order Autoregresssive Processes and Strong Mixing," Manuscript.

——— (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, I. (2017): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics (forthcoming)*, 1–27.

Andrews, I. and A. Mikusheva (2014): "Weak Identification in Maximum Likelihood: a Question of Information," *American Economic Review: Papers and Proceedings*, 104, 195–199.

——— (2015): "Maximum Likelihood Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 6, 123–152.

Bai, J. and S. Ng (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

Banerjee, A., M. Marcellino, and I. Masten (2016): "An Overview of the Factor-augmented Error-Correction Model," in *Advances in Econometrics*, ed. by E. Hillebrand and S. J. Koopman, Emerald Group, vol. 35, 3–41.

Barakchian, S. M. and C. Crowe (2013): "Monetary Policy Matters: Evidence from New Shocks Data," *Journal of Monetary Economics*, 60, 950–966.

Barndorff-Nielsen, O. E. and N. Shephard (2002): "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, 253–280.

Barro, R. J. . and G. Liao (2017): "Options-Pricing Formula with Disaster Risk," Manuscript, 1–48.

Benedikt M. Pötscher (1991): "Effects of Model Selection on Inference," *Econometric Theory*, 7, 163–185.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy: a Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120, 387–422.

Bharadwaj, P., L. Demanet, and A. Fournier (2017): "Deblending Random Seismic Sources via Independent Component Analysis," Manuscript.

Blanchard, O. and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655–673.

Blanco, D. and B. Mulgrew (2005): "ICA in Signals with Multiplicative Noise," *IEEE Transactions on Signal Processing*, 53, 2648–2657.

Blanco, D., B. Mulgrew, D. Ruiz, and M. Carrion (2007): "Independent Component Analysis in Signals with Multiplicative Noise Using Fourth-Order Statistics," *Signal Processing*, 87, 1917–1932.

Boivin, J., M. T. Kiley, and F. S. Mishkin (2011): "How Has the Monetary Transmission Mechanism Evolved over Time?" in *Handbook of Monetary Economics*, ed. by B. Friedman and M. Woodford, Elsevier, vol. 3A, 369–422.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T. and J. M. Wooldridge (1992): "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews*, 11, 143–172.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley (2017): "An Intertemporal CAPM with Stochastic Volatility," *Journal of Financial Economics (forthcoming)*.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?" in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, vol. 1, 65–148.

——— (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Coibion, O. (2012): "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics*, 4, 1–32.

Craine, R. and V. L. Martin (2008): "International Monetary Policy Surprise Spillovers," *Journal of International Economics*, 75, 180–196.

Cressie, N. A. C. (1993): *Statistics for Spatial Data*, New York: Wiley.

Dahlhaus, R. (2012): "Locally Stationary Processes," in *Time Series Analysis: Methods and Applications, Vol. 30*, ed. by T. Subba Rao, S. Subba Rao, and C. R. Rao, Oxford: Elsevier, chap. 13, 351–414.

Dalla, V., L. Giraitis, and P. C. Phillips (2015): "Testing Mean Stability of Heteroskedastic Time Series," Manuscript.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford: Oxford University Press.

Davis, R. A. and T. Mikosch (2009): "Extremes of Stochastic Volatility Models," in *Handbook of Financial Time Series*, ed. by T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen, Berlin: Springer, 355–364.

Diebold, F. X. and J. A. Lopez (1995): "Modeling Volatillity Dynamics," in *Macroeconometrics: Developments, Tensions and Prospects*, ed. by K. D. Hoover, New York: Springer, 427–472.

Douc, R., E. Moulines, and David S. Stoffer (2014): *Nonlinear Time Series: Theory, Methods and Applications with R Examples*, Boca Raton: CRC Press.

Doz, C., E. Renault, R. Chabot, and G. Thornton (2004): "Conditionally Heteroskedastic Factor Models: Identification and Instrumental Variables Estimation," Manuscript.

Eggertsson, G. B. and P. Woodford, Michael (2003): "Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 2003, 139–233.

Ehrmann, M. and M. Fratzscher (2017): "Euro Area Government Bonds - Fragmentation and Contagion During the Sovereign Debt Crisis," *Journal of International Money and Finance*, 70, 26–44.

Ehrmann, M., M. Fratzscher, and R. Rigobon (2010): "Stocks, Bonds, Money Markets and Exchange Rates: Measuring International Financial Transmission," *Finance and Development*, 47, 36–37.

Eichengreen, B. and U. Panizza (2016): "A Surplus of Ambition: Can Europe Rely on Large Primary Surpluses to Solve its Debt Problem?" *Economic Policy*, 31, 5–49.

Engle, R. (2001): "American Economic Association GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," 15, 157–168.

Feenstra, R. C. and D. E. Weinstein (2017): "Globalization, Markups, and US Welfare," *Journal of Political Economy*, 125, 1040–1074.

Fernandez-Perez, A., B. Frijns, and A. Tourani-Rad (2016): "Contemporaneous Interactions among Fuel, Biofuel and Agricultural Commodities," *Energy Economics*, 58, 1–10.

Fernandez-Villaverde, J. and J. F. Rubio-Ramirez (2007): "Estimating Macroeconomic Model: A Likelihood Approach," *Review of Economic Studies*, 74, .1059–1087.

Fernandez-Villaverde, J., J. F. Rubio-Ramirez, and M. S. Santos (2006): "Convergence Properties of the Likelihood of Computed Dynamic Models," *Econometrica*, 74, 93–119.

Flury, T. and N. Shephard (2011): "Bayesian Inference Based Only on Simulated Likelihood : Particle Filter Analysis of Dynamic Economic Models," *Econometric Theory*, 27, 933–956.

Foster, D. P. and D. B. Nelson (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 139–174.

Fuh, C.-D. (2006): "Efficient Likelihood Estimation in State Space Models," *Annals of Statistics*, 34, 2026–2068.

Gertler, M. and P. Karadi (2003): "Monetary Policy Surprises, Credit Costs and Economic Activity," *Brookings Papers on Economic Activity*, 34, 139–235.

Geweke, J. and R. Meese (1981): "Estimating Regression Models of Finite but Unknown Order," *International Economic Review*, 22, 55–70.

Gong, X., S. Yang, and M. Zhang (2017): "Not Only Health: Environmental Pollution Disasters and Political Trust," *Sustainability*, 9, 1–28.

Gospodinov, N. and S. Ng (2013): "Commodity Prices, Convenience yields, and Inflation," *The Review of Economics and Statistics*, 95, 206–219.

Gouriéroux, C. and A. Monfort (2014): "Revisiting Identification and Estimation in Structural VARMA Models," Manuscript.

Gurkaynak, R. S., B. P. Sack, and E. T. Swanson (2005): "Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements," *International Journal of Central Banking*, 1, 55–93.

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The Elements of Statistical Learning*, New York: Springer.

Hébert, B. and J. Schreger (2017): "The Costs of Sovereign Default: Evidence from Argentina," *American Economic Review (forthcoming)*.

Hogan, V. and R. Rigobon (2009): "Using Unobserved Supply Shocks to Estimate the Returns to Education," Manuscript.

Hyvärinen, A., K. Zhang, S. Shimizu, and H. P. O. (2010): "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity," *Journal of Machine Learning*

Research, 11, 1709–1731.

Islam, A., F. Islam, and C. Nguyen (2017): "Skilled Immigration, Innovation, and the Wages of Native-Born Americans," *Industrial Relations*, 56, 459–488.

Jacod, J. and P. Protter (2012): *Discretization of Processes*, New York: Springer.

Jahn, E. and E. Weber (2016): "Identifying the Substitution Effect of Temporary Agency Employment," *Macroeconomic Dynamics*, 20, 1264–1281.

Khalid, U. (2016): "The Effect of Trade and Political Institutions on Economic Institutions," *Journal of International Trade and Economic Development*, 26, 89–110.

Kilian, L. (1998): "Small-sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80, 218–230.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361–393.

Klein, R. and F. Vella (2009): "Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments," *Journal of Human Resources*, 44, 1047–1065.

Lanne, M. and H. Lütkepohl (2008): "Identifying Monetary Policy Shocks via Changes in Volatility," *Journal of Money, Credit and Banking*, 40, 1131–1149.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010): "Structural Vector Autoregressions with Markov Switching," *Journal of Economic Dynamics and Control*, 34, 121–131.

Lanne, M., M. Meitz, and P. Saikkonen (2017): "Identification and Estimation of Non-Gaussian Structural Vector Autoregressions," *Journal of Econometrics*, 196, 288–304.

Lazarus, E., D. J. Lewis, and J. H. Stock (2017): "The Size-Power Tradeoff in HAR Inference," Manuscript.

Lee, S.-W. and B. E. Hansen (1994): "Asymptotic Theory for the Garch(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Lewandowski, D., D. Kurowicka, and H. Joe (2009): "Generating Random Correlation Matrices Based on Vines and Extended Onion Method," *Journal of Multivariate Analysis*, 100, 1989–2001.

Lewis, D. J. (2017): "Robust Inference in Identification via Heteroskedasticity," Manuscript.

Lin, F., E. O. Weldemicael, and X. Wang (2016): "Export Sophistication Increases Income in Sub-Saharan Africa: Evidence from 1981-2000," *Empirical Economics*, 52, 1627–1649.

Lindner, A. M. (2009): "Stationarity, Mixing, Distributional Properties and Moments of GARCH(p,q)-Processes," in *Handbook of Financial Time Series*, ed. by T. Mikosch,

J.-P. Kreiß, R. A. Davis, and T. G. Andersen, Berlin: Springer, 43–69.

Ludvigson, S. C., S. Ng, and S. Ma (2016): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" Manuscript.

Lütkepohl, H. (2006): *New Introduction to Multiple Time Series Analysis*, Berlin: Springer.

Lütkepohl, H. and G. Milunovich (2016): "Testing for Identification in SVAR-GARCH Models," *Journal of Economic Dynamics and Control*, 73, 241–258.

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

McKay, A., E. Nakamura, and J. Steinsson (2016): "The Power of Forward Guidance Revisited," *American Economic Review*, 106, 3133–3158.

Millimet, D. L. and J. Roy (2016): "Empirical Tests of the Pollution Haven Hypothesis when Environmental Regulation Is Endogenous," *Journal of Applied Econometrics*, 31, 652–677.

Milunovich, G. and M. Yang (2013): "On Identifying Structural VAR Models via ARCH Effects," *Journal of Time Series Econometrics*, 5, 117–131.

Moessner, R. and W. R. Nelson (2008): "Central Bank Policy Rate Guidance and Financial Market Functioning," *International Journal of Central Banking*, 4, 204–226.

Mönkediek, B. and H. A. Bras (2016): "The Interplay of Family Systems, Social Networks and Fertility in Europe Cohorts Born Between 1920 and 1960," *Economic History of Developing Regions*, 31, 136–166.

Müller, U. K. (2012): "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models," *Journal of Monetary Economics*, 59, 581–597.

——— (2013): "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, 81, 1805–1849.

Nakamura, E. and J. Steinsson (2017): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics (forthcoming)*.

Normandin, M. and L. Phaneuf (2004): "Monetary Policy Shocks: Testing Identification Conditions under Time-Varying Conditional Volatility," *Journal of Monetary Economics*, 51, 1217–1243.

Olea, J. L. M., J. H. Stock, and M. W. Watson (2016): "Inference in Structural VARs with External Instruments," Manuscript.

Oppenheim, A. V. and G. C. Verghese (2010): "Signals, Systems, and Inference," Manuscript.

Pakel, C., N. Shephard, and K. Sheppard (2011): "Nuisance Parameters, Composite

Likelihoods and a Panel of Garch Models," *Statistica Sinica*, 21, 307–329.

Pavlova, A. and R. Rigobon (2007): "Asset Prices and Exchange Rates," *The Review of Financial Studies*, 20, 1139–1180.

Plagborg-Møller, M. (2017): "Bayesian Inference on Structural Impulse Response Functions," Manuscript.

Ramey, V. A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2A, 71–162.

Rigobon, R. (2003): "Identification through Heteroskedasticity," *The Review of Economics and Statistics*, 85, 777–792.

Rigobon, R. and D. Rodrik (2005): "Rule of Law, Democracy, Openness, and Income," *Economics of Transition*, 13, 533–564.

Rigobon, R. and B. Sack (2003): "Measuring the Reaction of Monetary Policy to the Stock Market," *Quarterly Journal of Economics*, 118, 639–669.

——— (2004): "The Impact of Monetary Policy on Asset Prices," *Journal of Monetary Economics*, 51, 1553–1575.

Romer, C. D. and D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

Sentana, E. and G. Fiorentini (2001): "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics*, 102, 143–164.

Shephard, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models: in Econometrics, Finance and Other Fields*, ed. by D. Cox, D. Hinkley, and O. Barndorff-Nielsen, New York: Chapman & Hall, 1–68.

Silverman, B. W. (1990): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

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

——— (2014): "Using Financial Data in Macroeconomic Models," in *Nonlinearities in Macroeconomics and Finance in the Light of Crises*, Frankfurt: ECB.

Sims, C., J. Stock, and M. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

Smets, F. and R. Wouters (2007): "Shocks and Frictions in US Business Cycle: a Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Stock, J. H. (2008): "Recent Developments in Structural VAR Modeling," in *NBER Summer Institute: Methods Lectures*.

Stock, J. H. and M. W. Watson (2002): "Macroeconomic Using Forecasting Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

——— (2016): "Factor Models and Structural Vector Autoregressions in Macroeconomics,"

in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Oxford: Elsevier, vol. 2A, 1–111.

——— (2017): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics," Manuscript.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference in Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, 80–108.

Swanson, E. (2016): "Measuring the Effects of Unconventional Monetary Policy on Asset Prices," Manuscript.

Trefethen, L. N. and D. Bau (1997): *Numerical Linear Algebra*, Philadelphia: Society for Industrial and Applied Mathematics.

Uhlig, H. (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Wright, J. H. (2011): "Term Premiums and Inflation Uncertainty: Empirical Evidence from an International Dataset," *American Economic Reivew*, 101, 1514–1534.

Zaefarian, G., V. Kadile, S. C. Henneberg, and A. Leischnig (2017): "Endogeneity Bias in Marketing Research: Problem, Causes and Remedies," *Industrial Marketing Management*, 65, 39–46.

A Proofs

A.1 Derivation of Proposition 1

Proof. I start with

$$E_t\left[\zeta_t \mid \sigma_t, \mathcal{F}_{t-1}\right] = L\left(H \otimes H\right) G\sigma_t^2.$$

Since v_t was shown to be a martingale difference sequence and $\mathrm{Var}_t\left(v_t\right) < \infty$ (B.2),

$$\mathrm{Cov}_t\left(v_t, v_s\right) = 0, \;\; s \neq t.$$

This also implies that in the signal-noise decomposition, (6), v_t is white noise. Using this fact, B.1-2, and the decomposition of ζ_t above, it is immediate that, for $s \neq t$,

$$E_t\left(\zeta_t\zeta_s'\right) = L\left(H\otimes H\right)GE_t\left[\sigma_t^2\sigma_s^{2'}\right]G'\left(H\otimes H\right)'L' \tag{13}$$
$$+ L\left(H\otimes H\right)GE_t\left[\sigma_t^2 v_s'\right] + E_t\left[v_t\sigma_s^{2'}\right]G'\left(H\otimes H\right)'L'.$$

By the law of iterated expectations, A.1 implies that

$$E_t\left[\Sigma_t \mid \sigma_s^2\right] = E_t\left[\varepsilon_t\varepsilon_t' \mid \sigma_s^2\right],\ t\geq s.$$

This, in turn, by the law of iterated expectations, implies that

$$E_t\left[vec\left(\varepsilon_t\varepsilon_t' - \Sigma_t\right)\sigma_s^{2'}\right] = 0,\ t\geq s.$$

Thus, setting $t > s$, the third term in (13) vanishes, leaving

$$E_t\left(\zeta_t\zeta_s'\right) = L\left(H\otimes H\right)GE_t\left[\sigma_t^2\sigma_s^{2'}\right]G'\left(H\otimes H\right)'L' + L\left(H\otimes H\right)GE_t\left[\sigma_t^2 v_s'\right]. \tag{14}$$

Finally, I can rewrite (14) as

$$L\left(H\otimes H\right)\left(GE_t\left[\sigma_t^2\sigma_s^{2'}\right]G' + GE_t\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)\right]\right)\left(H\otimes H\right)'L'$$
$$= L\left(H\otimes H\right)G\tilde{M}_{t,s}\left(H\otimes H\right)'L' \tag{15}$$

where $\tilde{M}_{t,s} = E_t\left[\sigma_t^2\sigma_s^{2'}\right]G' + E_t\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)'\right]$. $\tilde{M}_{t,s}$ is an $n\times n^2$ matrix. Proposition 1 then follows directly. □

A.2 Derivation of Proposition 2

Proof. It is necessary to show that given Assumption C, $E_t\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)'\right] = E_t\left[\sigma_t^2 u_s'\right]G'$ where $u_s = matdiag\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)$. Assumption C states that $E_t\left[\sigma_{it}^2\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)\right]$ is diagonal for all $i = 1, 2, \ldots, n$. Therefore, $E_t\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)'\right]$ has columns of zeros except for those pertaining to a diagonal element of $\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)'$, i.e. for $j = 1, 2, \ldots, n$, column $j + (j-1)n$ is equal to $E_t\left[\sigma_t^2\left(\varepsilon_{js}^2 - \sigma_{js}^2\right)\right]$. Now by the definition of G, AG' takes the jth column of the $n\times n$ matrix A , $j = 1, 2, \ldots, n$, and places it in column $j + (j-1)n$ of a matrix of zeros. Therefore, if the jth column of A is equal to $E_t\left[\sigma_t^2\left(\varepsilon_{js}^2 - \sigma_{js}^2\right)\right]$ for all $j = 1, 2, \ldots, n$, then $AG' = E_t\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)'\right]$. This is true if $A = E_t\left[\sigma_t^2 diag\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)\right] = E_t\left[\sigma_t^2 u_s'\right]$. Thus $E\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)'\right] = E\left[\sigma_t^2 u_s'\right]G'$ as desired. Proposition 2 follows from re-writing the two terms of (13) in this way, and summing to yield $\check{M}_{t,s}$. □

A.3 Proof of Theorem 1

I begin by proving two lemmas for properties of the singular value decomposition (SVD).

Definition 1. Define

- $U_1 D_U U_2' = V$, a reduced SVD, V $n_1 \times n_2$, D_U $d \times d$,
- C_i is a full rank matrix, $m_i \times n_i, m_i \geq n_i$,
- $F = C_1 V C_2'$, non-defective.

Lemma 1. *There exists a matrix Γ_1 such that $CU_1\Gamma_1$ is an orthogonal matrix of singular vectors from a SVD of F.*

Proof. Define $Q_1 R_1 = CU_1$, a reduced QR decomposition, and similarly for CU_2. Then $F = Q_1 R_1 D_U R_2' Q_2'$. R_1 is $d \times d$ and full rank since, given CU_1 is full rank d, it has no zeros on the diagonal (Trefethen & Bau (1997), Exercise 7.5). Now define $W_1 D_R W_2' = R_1 D_U R_2'$, another SVD; then $F = (Q_1 W_1) D_R (W_2' Q_2')$ is a reduced SVD (it is easily shown D_R are singular values of F, and the corresponding vectors are clearly orthogonal). Thus write $Q_1 R_1 \left(R_1^{-1} W_1\right) = Q_1 W_1$ so $\Gamma_1 = R_1^{-1} W_1$, which is guaranteed to exist. □

Definition 2. Define

- $S_1 D_S S_2' = F$, a reduced SVD

Lemma 2. *The SVD of F is unique up to rotations characterized by $F = S_1 T_1 D_S T_2 S_2'$ where T_i is orthogonal*

Proof. The singular values, D_S, are unique, singular vectors corresponding to non-repeated values are unique up to sign, and the space of vectors corresponding to k repeated singular values corresponds to linear combinations of any k such vectors. Thus $F = (S_1 T_1) D_S (T_2 S_2')$ characterizes any reduced SVD as T_i can incorporate any such sign changes or linear combinations. Since $S_i T_i$ must be orthogonal, $T_i' S_i' S_i T_i = I_d$. Then since S_i is orthogonal, $T_i' T_i = I_d$, so T_i is orthogonal. □

Definition 3. Define

- In particular, $C_1 = (H \otimes H) G$, $n^2 \times n$ with rank n,
- G is a selection matrix such that $vec\,(ADA') = (A \otimes A)\, G diag\,(D)$,
- $\hat{S}_1 = C_1 U_1 \Gamma_1 T_1$, an arbitrary reduced SVD of F,

- V is $n \times n^2$ and has no scalar multiple rows,
- $rank\,(V) \geq 2$.

Proposition 3. *H is uniquely determined from the equations $F = C_1 V C_2'$ provided V has no scalar multiple rows.*

Proof. U_1 is $n \times d$. Note $CU_1 = \left[vec\left(Hdiag\left(U_1^{(1)}\right)H'\right), \ldots, vec\left(Hdiag\left(U_1^{(d)}\right)H'\right)\right]$, where $d \geq 2$. By the scalar multiples condition on V, for any column j of H, there exists at least one pair k, l such that $U_{1,j}^{(l)}/U_{1,i}^{(l)} \neq U_{1,j}^{(k)}/U_{1,i}^{(k)}$ for all $i = 1, 2, ..., d,\ i \neq j$. By an argument due to Sims (2014), $H^{(j)}$ is unique up to scale and sign as the right eigenvector of $Hdiag\left(U_1^{(l)}\right)H'\left(Hdiag\left(U_1^{(k)}\right)H'\right)^{-1}$ corresponding to the jth eigenvalue. The same argument applies to $C\tilde{U}_1$ where $\tilde{U}_1 = U_1 \Gamma_1 T_1$, provided $\tilde{U}_1$ has no scalar multiple rows. To establish this, take any two rows in U_1 that are not scalar multiples; multiplying by full-rank Γ_1 cannot decrease their rank (so they do not become scalar multiples). The same holds true for multiplication by the orthogonal T_1. Thus H remains the unique solution to $C\tilde{U}_1$. □

Proposition 3 is re-written in economic terms to yield Theorem 1.

A.4 Proof of Corollary 1

Proof. Corollary 1 follows directly from Proposition 3 above for any column j for which a pair k, l exists such that $U_{1,j}^{(l)}/U_{1,i}^{(l)} \neq U_{1,j}^{(k)}/U_{1,i}^{(k)}$ for all $i = 1, 2, ..., d$. □

A.5 Proof of Theorem 2

Proof. Theorem 2 is based on the argument underlying Proposition 2. Note that the vectorization of $E_t\left[\zeta_t\right]$ is given by $vech\left(HE_t\left[\Sigma_t\right]H'\right)$, an additional equation of the form found in CU_1. Define $U_{1,M} D_M U_{2,M}' = M_{t,s}$ and $\hat{M} = \begin{bmatrix} U_{1,M} & E_t\left[\sigma_t^2\right] \end{bmatrix}$. Then there is an additional column over which to find a k, l pair for j such that $\hat{M}_j^{(l)}/\hat{M}_i^{(l)} \neq \hat{M}_j^{(k)}/\hat{M}_i^{(k)}$ for all $i = 1, 2, ..., d\ i \neq j$. The condition on $M_{t,s}$ (V in Proposition 3) guaranteeing this is augmented to require no scalar multiple rows in $\begin{bmatrix} M_{t,s} & E_t\left[\sigma_t^2\right] \end{bmatrix}$. Note that this logic can be extended to adding additional autocovariances, etc., in each case making the length of the rows that must not be scalar multiples longer and thus decreasing the plausibility of the condition failing. □

A.6 Proof of Corollary 2

Proof. The result is immediate given that, if H is identified, a moment of ε_t, $g(\varepsilon_t)$, is identified as $g(H^{-1}\eta_t)$. By the condition of the corollary, these moments are sufficient to identify the parameters θ. □

Remark. The condition of the corollary is not particularly restrictive. If conditional heteroskedasticity is not present, obtaining moments is further simplified as autocovariances of ζ_t do not depend on past shock values, so

$$Cov\left(\sigma_t^2, \sigma_s^2\right) = E_t\left[vec\left(H^{-1}\eta_t\eta_t'\left(H'\right)^{-1}\right) vec\left(H^{-1}\eta_s\eta_s'\left(H'\right)^{-1}\right)'\right].$$

Depending on the functional form, a normality (or similar) assumption on ε_t may be required to back out the variance of ζ_t, although obtaining the variance may not be necessary for identification of the underlying parameters. GARCH parameters are known to be identified from moments of ε_t. It is a matter of algebra to show that the same is true for other processes, such as the AR(1) log SV process. For example, under the assumption of stationarity and normal innovations, properties of the lognormal distribution dictate that the autoregressive parameter, and thus the innovation variance, is identified from the first and second autocovariance of σ_t^2, dimension by dimension. From there, it is possible to back out means and innovation covariances using additional moments.

B Methods

B.1 Weak Identification

Any discussion of identification must be tempered by the possibility of weak identification when estimation may be based on small samples. While there is now a clear understanding of the problems posed by weak identification, beginning with Stock & Wright (2000), there remains much work to be done to develop methods in more complex settings in terms of both testing for identification and conducting robust inference. For example, Andrews (2017) presents possible the first comprehensive approach suited to GMM.

It is important to note potential sources of weak identification in this setting. Naturally, near-zero variation in the volatilities destroys the identifying variation that this scheme seeks to exploit. This is the leading case of weak identification. Similarly, if the variation in volatilities is small relative to that of the i.i.d. disturbances each period, identification will be difficult in small samples. The other threat is from the "scalar multiples" problem

highlighted in Theorems 1 and 2. While this seems unlikely to hold in population, it may be the case that the differences are negligible and hard to estimate in small samples. If two shocks' variances follow a common persistent factor (with some transient noise), the autocovariance matrix would have two rows very close to proportional to each other. This would lead to a breakdown of the identifying argument.

It is easy enough to test whether these pathologies exist using conventional methods. For example, it is possible to test the null hypothesis that the autocovariance matrix of σ_t^2 is equal to zero at an arbitrary level of significance, or that its rows are related by scalar multiples. However, the challenge in this literature has been to assess what the appropriate critical values are in each setting for such tests. What is the mapping between the size of a test on such parameters and the bias of the estimates of interest, or the size distortion of tests on those parameters? This is what Stock & Yogo (2002) accomplish for IV, but it is generally an open question.

The methodological frontier can be assessed by estimation approach. Andrews (2017) provides a 2-step method to assess the strength of identification in a GMM setting. Essentially, this involves computing a robust confidence set and a strong identification confidence set, and comparing them. Conventional weak identification analysis is based on the CUE estimator, largely because this permits the use of convenient χ^2 critical values. However, given the recency of Andrews' work, no studies have yet applied it in a numerically challenging high-dimensional setting. As discussed in Section B.2, the CUE estimator is frequently unstable in this highly non-linear context, with the weighting matrix frequently near-singular. Since the construction of a robust set generally requires repeated optimization over a fine grid, this is a substantial problem. Methods to compute alternative critical values based on GMM estimators using other alternative, simpler weighting matrices can mitigate these problems. However, further work is required to establish the conditions under which conclusions based on such alternative tests can be extrapolated back to estimators using more complex weighting matrices. As an econometric problem in its own right, this is outside the scope of the current paper. Because the difficulty in applying the Andrews methods arises in the construction of the robust set, it also rules out robust inference on estimated parameters for the time being.

In the maximum likelihood context, Andrews & Mikusheva (2015) offer a method to assess strength of identification via the deviation of two alternative estimates of the Fisher information. However, the current setting is again more complicated than those considered in that paper. Moreover, instead of actual maximum likelihood, here it is approximated via MCMC. This means that the theoretical results of Andrews & Mikusheva (2015) cannot be trivially extended.

Work is even more nascent in the Bayesian context for assessing the strength of identification. The most relevant work is Müller (2012), but again, this provides less insight into the MCMC methods I must adopt.

In sum, weak identification tests and robust inference remain an unsolved problem in this setting. This is regrettable, as in proposing a new method of identification, it is important to assess the strength of the identifying variation it exploits in practice. With future work, this will hopefully become possible. For now, results are carefully compared across estimation strategies in an effort to gauge, in the absence of strong identification, to what extent minor functional form assumptions or particular estimation approaches shape the findings. Encouragingly, evidence from the simulation study, calibrated to the empirical application, suggests that identification is strong, even in a shorter sample (and, to some extent, even when the identifying variation is highly reduced).

B.2 Additional estimation methods

B.2.1 Practical points on quasi-likelihood based inference

As noted in the text, simulation evidence recommends the use of quasi-likelihood based inference in this setting. The drawback of any likelihood-based approach is the necessity of specifying a law of motion for the structural variance. Unfortunately, since the variance path is unobserved, evaluating the likelihood is a problem requiring difficult integration in each time period over all possible past values of σ_t^2. Based on the chosen likelihood function, numerical integration allows the model to be estimated via Markov Chain Monte Carlo (MCMC).[20]

For the purposes of this paper, Hamiltonian MCMC is adopted, implemented via the MatlabStan package. The Stan software offers interfaces for virtually all commonly-used statistical packages, and is highly recommended for applied work. An advantage of Stan is that it chooses the tuning parameters of the MCMC adaptively during the warm-up period, increasing the odds of obtaining well-behaved chains. This means that the researcher need only supply Stan with a file describing the model and parameters to be estimated and the data. Sample files for the models used in this paper will be made available on my website.

A researcher should, in general, consider multiple chains to avoid local minima (since each chain is random, they take different paths). However, different chains may converge to

[20]Maximum likelihood can be well approximated by MCMC methods, as discussed in Flury & Shephard (2011). This is the case when MCMC has flat priors on the parameters and the posterior is concentrated around its mode. Convergence results for this methodology can be found in e.g. Fernandez-Villaverde, Rubio-Ramirez, & Santos (2006), Fernandez-Villaverde & Rubio-Ramirez (2007), Ackerberg, Geweke, & Hahn (2009), and Douc, Moulines, & Stoffer (2014) (Sections 2-3), amongst others.

different parameter values that are observationally equivalent, but correspond to different labelings of the columns of H. I recommend labeling the shock series estimated by each chain separately before combining the chains to compute overall estimates, to avoid averaging over the columns of H. Some experimentation is necessary, depending on the dimension and distribution of data, to determine how many iterations are needed for a chain to converge to its stationary distribution. The main results in my empirical application are based on four chains of 10,000 iterations, each with a 2000-iteration warm-up. If the researcher has priors over the model parameters, these can be incorporated to perform Bayesian inference on the parameters of interest. These can easily be appended to a Stan model file. As with any Bayesian estimation problem, care must be taken to choose priors with suitable properties.[21]

These methods have the advantage of directly computing a distribution of values for the filtered volatility path, which are a potentially interesting object in themselves. For example, plotting the paths of the volatility of monetary policy shocks and inflation over time can suggest whether periods of economic turmoil were driven by dramatic movements in inflation or systematically precarious policy-making, in a way that the previous custom of simply looking at implied shocks (with constant volatility) cannot. The downside is the additional burden on the researcher to specify a functional form.

The asymptotic properties of MCMC estimation remain difficult to derive in more complicated models. As such, no specific results have been derived for models having the complexity displayed here (although they have for univariate SV models). General discussion of such asymptotic theory can be found in Douc, Moulines & Stoffer (2014), Sections 2-3, Fernandez-Villaverde, Rubio-Ramirez, & Santos (2006), and Fernandez-Villaverde & Rubio-Ramirez (2007).

B.2.2 GMM

While GMM (or minimum distance) seems a natural choice to implement TVV-ID, its merits must be carefully weighed. It has the advantage of being entirely non-parametric. It can be applied without making any further assumptions on the process σ_t^2; the matrix $M_{t,s}$ is estimated as a nuisance matrix. GMM in this context can be built around an autocovariance (likely the first) which, as argued above, absent some specific deficiencies, is sufficient to provide identification. However, estimation can be improved by exploiting additional moments. For instance, the mean of ζ_t, the covariance $E_t\left[\eta_t\eta_t'\right]$, contains much information, even if that

[21] In contexts where some off-diagonal elements of the H matrix are thought to be small, a moderately strong prior has the additional advantage of making it more likely that multiple chains result in the same shock ordering. This occurs because an ordering placing a relatively small element on the diagonal (which is normalized to unity) will inflate the off-diagonal elements of that column, which is discouraged by the prior.

information alone could not identify H.[22] Also recall that additional moments reduce the possibility of weak or non-identification, as discussed in Theorem 3. For standard GMM asymptotic results to apply, the σ_t^2 process must additionally be second-order stationary.

GMM presents a difficult high-dimensional optimization problem. For example, a standard $3-$dimensional VAR leaves 27 parameters to estimate provided a single autocovariance and the mean of ζ_t are used (as is the case in simulations and unreported applications in this paper). The parameter space can be reduced by making additional assumptions on the volatility processes, but a key virtue of TVV-ID (and a GMM implementation in particular) is avoiding such assumptions. Given the high dimension of this problem and the degree of non-linearity, the optimization can be numerically challenging with many highly pronounced local minima (as opposed to a flat objective function). Regardless of the optimization routine used, it is difficult to be certain a minimum is global, and results are highly dependent on start-values. This is particularly true given the numerical instability introduced with attractive forms of weighted GMM, for example the efficient continuously updated estimator (CUE). This estimator frequently involves the inversion of a nearly singular matrix on each iteration. If care is not taken, minimization can result in a negative value of the objective function by virtue of a non-positive definite weighting matrix. These numerical issues are the least appealing feature of GMM estimation in this setting.

Unlike in many other identification schemes common in SVARs, TVV-ID is highly overidentified; as such, it is possible to conduct standard misspecification tests on various aspects of the model. For example, a $J-$test could be used to detect instability in H. Another advantage of GMM is that it admits established parameter stability tests, like the sup-Wald etc. tests of Andrews (1993).

B.2.3 GARCH

With my new results demonstrating identification for any functional form (that implies an autocovariance), it is worth reconsidering the role GARCH can play. Since identification is no longer reliant on the GARCH assumption in a knife-edge sense, GARCH can be evaluated in terms of how well it describes the data (see e.g. Diebold & Lopez, 1995, Kim, Shephard, & Chib, 1998, or Barndorff-Nielsen & Shephard, 2002 for a comparison of functional forms) and its performance in simulation. Second, since it is now possible to identify and estimate a model using multiple functional forms (via the likelihood approaches discussed above), it is possible to directly evaluate whether assuming a restrictive form like GARCH has a significant impact on the results obtained. In practice, the GARCH model is easily estimated

[22] The same is true of the variance of ζ_t if a non-Gaussianity assumption is imposed, in keeping with Gouriéroux & Monfort (2014).

in this context using maximum likelihood as in Normandin & Phaneuf (2014), and others. Under the standard GARCH assumptions on parameters and distributions, the usual maximum likelihood asymptotics apply; this is also a trivial extension of the QML discussion in Section 4.2. In simulation, the GARCH estimator performs worse than the hybrid GARCH method in many settings. Primarily, this is due to substantial excess mass around zero in the distribution of estimators. It is likely that this results from the estimation routine being driven to local minima at the upper bound of the parameter space (the explosive region), forcing H estimates to zero. Hybrid GARCH does not suffer from this weakness due to the selection of those parameters via calibration.

B.2.4 Infill Asymptotics

Infill asymptotics examine the behaviour of estimators as the frequency of observation increases, as opposed to the length of time spanned by the observations. In an infill setting (see e.g. Cressie, 1993, Section 5.8 or Dahlhaus, 2012, Section 2 for an introduction), an expectation can be consistently estimated over a finite time span as observations are taken over an increasing density of intervals. Infill asymptotic arguments are well-suited to non-stationary time series, (where standard asymptotics do not hold), that can be approximated by a stationary process in some neighbourhood of each point in time. Foster & Nelson (1996), provide an application to covariance estimation. The discussion below is intended as a non-technical overview; the interested reader should consult the references noted.

Kernel method

If a noisy time series, like the reduced-form variances considered here, is locally stationary, kernel smoothing can help eliminate the noise by smoothing values across a small window, see e.g. Hastie, Tibshirani, & Friedman (2009), Chapter 6. The objective is to estimate the regression function $E_t\left[\eta_t\eta_t' \mid \eta_{1:T}\right]$ via

$$\overline{\eta_t\eta_t'} = \sum_{N_{b_T}(t)} R\left(\eta_\tau\eta_\tau'\right),$$

where $R\left(\cdot\right)$ is a weighting function. More concretely, for a symmetric kernel $k\left(l; b_T\right)$,

$$\overline{\eta_t\eta_t'} = \frac{1}{b_T}\sum_{l=-b_T}^{b_T} k\left(l; b_T\right)\eta_{t+l}\eta_{t+l}'$$

where b_T is the bandwidth (I follow the discussion of local covariance estimation in Dahlhaus, 2012, Section 2.2, with the exception that I subsume his $b_T T$ into simply b_T). The consensus

in the statistics literature is that the choice of kernel is relatively unimportant, driven by the high relative efficiency of many kernels relative to the "optimal" Epanechnikov kernel, see e.g. Silverman (1990), pg. 43. For the purposes of this paper, the Epanechnikov kernel is used, defined as

$$k_{EP}\left(l;b_T\right)=\begin{cases}\frac{3}{4}\left(1-\left(\frac{l}{b_T}\right)^2\right) & |l/b_T|\leq 1\\ 0 & \text{otherwise,}\end{cases}$$

see Hastie, Tibshirani, & Friedman (2009), pg. 193. The bandwidth used in the simulation study is $b_T = 12 = \lambda T$, corresponding to a window twelve months either side of an observation. Experimentation with the bandwidth on the empirically calibrated AR(1) SV DGP shows very similar performance for values of b_T ranging from 6 to 24, with the minimum mean square error occurring at 12. When the data is non-stationary, the choice of bandwidth will be constrained to ensure the neighbourhood of smoothing is locally stationary.

Applying a kernel smoothing algorithm estimates the path $\widehat{\eta_t\eta_t'}$ from $t = b_T+1$ to $T-b_T$. Proposition 3 of Sentana & Fiorentini (2001) shows that under very general conditions, H is identified from such a path. This presents a very high-dimensional overidentified minimum distance problem – $\left(n^2+n\right)/2$ parameters in H and $\left(T-2b_T\right)n$ structural variances.

The asymptotic properties of kernel estimates are discussed in detail in Theorem 3 in Dahlhaus (2012). To summarize, under regularity and strong smoothness conditions, Dahlhaus shows that estimates are asymptotically normal with a rate of $\left(\lambda T\right)^{-1/2}$. The estimates also exhibit bias depending on the degree of non-stationarity of the true DGP. Under weaker conditions, slower rates are possible. Dahlhaus' result is quite general and also covers the convergence of functions of the smoothed estimates, like the minimum distance problem considered here.

In the 2-dimensional case in this paper's simulations, it is possible to greatly reduce the dimensionality of the resulting minimum distance problem by avoiding the need to directly estimate the nuisance matrix, the two structural variances for each observation. The standard minimum distance set-up with three moments for the unique elements of $\overline{\eta_t\eta_t'}$ in each time period is asymptotically equivalent to minimum distance on the difference between the off-diagonal elements of $H^{-1}\overline{\eta_t\eta_t'}H'^{-1}$ and zero. Given the nature of the problem, first estimating a smoothed path and then performing highly non-linear minimum distance on that path, there do not appear to be established methods for inference in this setting; as such, standard errors are omitted for this estimator in my simulations.

Blocking method

A second approach retains the flavour of Rigobon's identification argument. Essentially, if at least local stationarity is assumed within intervals of fixed length, the mean of ζ_t can be estimated over blocks of data; the estimates will be consistent as Δt, the time increment, tends to zero. Then, based on these subsamples, H is estimated using some minimum distance formulation following Proposition 3 of Sentana & Fiorentini (2001).

This has the downside that it is sensitive to the sub-sample length, as suggested by unreported simulations, but it is not susceptible to the main criticism leveled at Rigobon's approach in Section 4, since arbitrarily-divided sub-samples of fixed length will not result in systematic non-diagonality. The expectation over each sub-sample remains fixed and generally differs across sub-samples even if the process is stationary, as the length of the sub-sample does not need to increase to apply the infill argument to estimators. Within each of these blocks, say $[\underline{t}, \bar{t})$, estimate $\int_{\underline{t}}^{\bar{t}} E_t \left[H\Sigma_t H'\right] dt$ via $\hat{\zeta}_{\underline{t}:\bar{t}} \equiv vech\left(\frac{1}{T}\sum_{t=\underline{t}}^{\bar{t}} H\Sigma_t H'\right)$. From this "path" of discrete segments, H can be estimated. If the process is in fact stationary, under various sets of assumptions, as detailed extensively in Jacod & Protter (2012), convergence to $E\left[H\Sigma_t H' \,|_{\underline{t}}^{\bar{t}}\right]$ occurs for each block of length λT at the standard rate $(\lambda T/\Delta t)^{-1/2}$. For example, if η_t can be approximated on $[\underline{t}, \bar{t})$ as an Itô semi-martingale (and some regularity conditions hold), their most basic CLT, Theorem 5.1.2, applies. If only local stationarity is assumed, the convergence results noted above for kernel estimators apply under the conditions noted in Dahlhaus (2012). As with the kernel method, it remains to estimate H from the overidentified system (assuming there are more than two blocks). Given these estimated innovation covariances for each block of data, a minimum distance estimator can be used to find the optimal H to satisfy the overidentified system of equations.

B.3 Least Mean Square Filter

As discussed in the text, the variance paths of the structural shocks may be an object of interest in their own right. Some estimation methods produce these directly; MCMC draws values for the paths in order to conduct its numerical integration, and GARCH-based methods imply a filtered path. For others, like GMM, the estimated nuisance matrix instead provides limited information on the moments of the variance process. In these cases, a filtering algorithm is needed to obtain estimates of the variance paths, ideally one that requires few additional assumptions.

The Wiener filter is non-parametric, based only on moments of the measurement and the signal. Since the predictions are linear, this filter of course can produce negative variance estimates. This discussion follows that of Oppenheim & Verghese (2010) Chapter 11, and

primarily considers the Finite Impulse Response (FIR) version. The reader should consult the reference for a more detailed treatment. The Wiener filter has two basic assumptions:
Assumption W.

1. *Signal $s_t$and noise v_t are stationary for all $t = 1, 2, \ldots, T$,*

2. *s_t and v_t have known autocorrelation and cross-correlation.*

For this discussion, I strengthen these conditions in Assumption W′ for simplicity.
Assumption W′. *For all $t = 1, 2, \ldots, T$:*

1. *The signal is σ_t^2, with $diag\left(\sigma_t^2\right) = E\left[\varepsilon_t \varepsilon_t' \mid \sigma_t\right]$,*

2. *The noise is $vec\left(\varepsilon_t \varepsilon_t' - \Sigma_t\right)$ (where $\Sigma_t = diag\left(\sigma_t^2\right)$),*

3. *σ_t^2 and $\left(\varepsilon_t \varepsilon_t' - \Sigma_t\right)$ are independent (a SV model),*

4. *ε_t is independently normally distributed with $E\left[\Sigma_t^{-1/2} \varepsilon_t \varepsilon_t' \Sigma_t^{-1/2}\right] = I_n$.*

Assumption W′.3 implies that W.2 can be satisfied based on the data, without *a priori* knowledge of the signal or noise's moments or training data on s_t. Assumption W′.4 will be exploited later. The Wiener filter minimizes the mean-squared error of the linear prediction

$$s_t = \sum_{j=0}^{J} a_j x_{t-j}$$

where x_t and s_t are vectors of the same length and a_j is a square matrix of the same dimension. J is the maximum lag considered in the prediction problem. Taking expectations and first order conditions results in the Wiener-Hopf equations,

$$\begin{bmatrix} R_x(0) & R_x(1) & \cdots & R_x(J) \\ R_x(1) & R_x(0) & \cdots & R_x(J-1) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(J) & R_x(J-1) & \cdots & R_x(0) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_J \end{bmatrix} = \begin{bmatrix} R_{xs}(0) \\ R_{xs}(1) \\ \vdots \\ R_{xs}(J) \end{bmatrix},$$

where $R_x(j) = E\left[x(t)\, x(t+j)'\right]$ and $R_{xs} = E\left[x(t)\, s(t+j)'\right]$, or, more compactly,

$$R_x = a R_{xs}.$$

The elements of R_x are easily estimated, as x_t is the input. However, it remains to calculate R_{xs}. Now it is necessary to apply the assumptions of W and W′. First, observe that in

general, the input data is $x_t = vech\left(\widehat{\varepsilon_t\varepsilon_t'}\right)$, where $\hat{\varepsilon}_t = \hat{H}^{-1}\eta_t$. However, the off-diagonal elements of $\varepsilon_t\varepsilon_t'$ are irrelevant for the linear prediction of both $\varepsilon_u\varepsilon_u', u \neq t$ and $\Sigma_u \forall u$. In other words,

$$E\left[\varepsilon_{kt}\varepsilon_{lt}\varepsilon_{iu}\varepsilon_{mu}\right] = 0, k \neq l, t \neq u$$

$$E\left[\varepsilon_{kt}\varepsilon_{lt}\Sigma_u\right] = 0, k \neq l, t \neq u$$

by W′.3 and W′.4. Therefore, specialize $x_t = diag\left(\widehat{\varepsilon_t\varepsilon_t'}\right)$, and estimate R_x using the obvious sample counterpart. Now, turn to the estimation of R_{xs}. By assumptions W′.3 and W′.4., $R_{xs}(j) = R_x(j)$ for $j = 1, 2, \ldots, J$. It remains to estimate $R_{xs}(0)$. Again under assumptions W′.3 and W′.4,

$$E\left[diag\left(\varepsilon_t\varepsilon_t'\right) diag\left(\varepsilon_t\varepsilon_t'\right)'\right] = \begin{bmatrix} 3E\left[\sigma_{1t}^4\right] & E\left[\sigma_{1t}^2\sigma_{2t}^2\right] & \cdots & E\left[\sigma_{1t}^2\sigma_{nt}^2\right] \\ E\left[\sigma_{2t}^2\sigma_{1t}^2\right] & 3E\left[\sigma_{2t}^4\right] & \cdots & E\left[\sigma_{2t}^2\sigma_{nt}^2\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[\sigma_{nt}^2\sigma_{1t}^2\right] & E\left[\sigma_{nt}^2\sigma_{2t}^2\right] & \cdots & 3E\left[\sigma_{nt}^4\right] \end{bmatrix}.$$

Thus, to obtain the desired matrix, $E\left[\sigma_t^2\sigma_t^{2\prime}\right]$, the diagonal of an estimator of the above matrix can simply be divided by one-third. Thus, estimators of both R_x and R_{xs} are easily available under assumptions W and W′. Finally, obtain $\hat{a} = \hat{R}_x^{-1}\hat{R}_{xs}$. This requires that $\hat{R}_x$ be invertible. Asymptotically, this will be the case since R_x is a covariance matrix. Then the filtered path for σ_t^2 is obtained as

$$\hat{\sigma_t^2} = \sum_{j=0}^{t-1} \hat{a}_j diag\left(\widehat{\varepsilon_t\varepsilon_t'}\right)_{t-j}.$$

While the above discussion is highly specialized given assumption W′, it is possible to relax W′ and obtain similar results using more algebra.

B.4 Standard errors

Each of the estimation schemes proposed in the text comes with its own method to compute standard errors for the H matrix and other parameters. However, in general, the innovations on which this analysis is based have to be computed based on data, as in a VAR. Given this fact, it is worthwhile discussing the construction of standard errors.

Block diagonality of the variance-covariance matrix

For SVARs, it is a familiar result that the asymptotic covariance matrix of moments for the autoregressive coefficients and the covariance decomposition has a block diagonal structure, see e.g. Lütkepohl (2006). This extends to reduced-form IRF coefficients, which also have a block diagonal structure with respect to the covariance decomposition block. This follows from a delta method argument, since the reduced-form IRF coefficients are functions of just the autoregressive coefficients. As this can greatly simplify the computation of standard errors, it is important to verify that this result still holds for identification arguments based on higher moments, as is the case here. Throughout the discussion, η_t is assumed to come from a symmetric distribution with mean zero; this is necessary considering the use of higher moments. First, consider GMM (with the additional assumption of second-order stationarity). The off-diagonal blocks take the form

$$\begin{aligned}
&E\left[vec\left(\eta_u Y_{u-j}'\right)\left(vec\left(cov\left(vech\left(\eta_t\eta_t'\right), vech\left(\eta_s\eta_s'\right)\right)\right) - L\left(H\otimes H\right)GM_{t-s}\left(H\otimes H\right)' L'\right)'\right] \\
&= E\left[vec\left(\eta_u Y_{u-j}'\right) vec\left(cov\left(vech\left(\eta_t\eta_t'\right), vech\left(\eta_s\eta_s'\right)\right)\right)'\right] \\
&\quad - E\left[vec\left(\eta_u Y_{u-j}'\right)\left(L\left(H\otimes H\right)GM_{t-s}\left(H\otimes H\right)' L'\right)'\right] \\
&= 0
\end{aligned}$$

In other words, the lower block is a minimum distance problem, and, in expectation, the value of any population moments are uninformative for η_u beyond its mean-zero property: $E\left[\eta_u \mid \eta_t\eta_t', Y_{u-j}\right] = 0, j = 1, 2, \ldots, J$.

This remains the case if the second stage – the covariance decomposition – is obtained via a log-likelihood instead of moments. It is required that

$$E\left[vec\left(\eta_t Y_{t-j}'\right)\frac{\partial \log f\left(\eta_t \mid \theta\right)'}{\partial\theta}\right] = 0$$

for all $t = 1, 2, \ldots, T$ and lags $j = 1, 2, \ldots, J$, where $f\left(\eta_t \mid \theta\right)$ is a likelihood. If $f\left(\eta_t \mid \theta\right)$ depends on η_t only through even moments of η_t (as in a multivariate normal), and η_t is symmetric, then the result follows from $E\left[\eta_u \mid \eta_t\eta_t', Y_{u-j}\right] = 0, j = 1, 2, \ldots, J$, and similar conditions if the distribution involves higher moments. The same holds if the log-likelihood $f\left(\eta_t \mid \theta\right)$ is replaced by a log-posterior $f\left(\theta \mid \eta_t\right)$ with the same forms of dependence on η_t. The block-diagonal structure is then carried forward into any IRF variance-covariance matrix.

MCMC draws from an approximation to the posterior $f\left(\theta \mid \eta_t\right)$. As argued above, in the exact maximization of the posterior, the covariance of the structural parameter estimates

with the reduced-form parameter estimates is block diagonal. If MCMC is carefully constructed so as to yield a good approximation to the posterior, then the covariance of the MCMC draws of θ with the reduced-form parameters will also be approximately zero. The IRF covariance can be constructed from a block diagonal matrix formed from the covariance of the reduced-form VAR coefficients, and the covariance obtained for H via MCMC.

Bootstrapping for factor estimation

The empirical application considered here adds a further complication. Since the factors used are generated regressors, it is necessary to compute the covariance matrix for the autoregressive parameters via a bootstrap procedure. The block diagonal structure discussed above means that draws of the reduced-form parameters and draws of the H parameters can be obtained independently. The method selected, designed for such a setting, is closely based on the block bootstrap of Gospodinov & Ng (2013). This method draws from the true data T/λ_T blocks of length λ_T. Without imposing a parametric model, this preserves much of the lag structure and interdependence of innovations present in such time series data. Experimentation shows little dependence on the block length. In the application, $\lambda_T = 14$, in keeping with the lag structure assumed in the estimation itself (13 lags) by BBE and subsequently here. This procedure is used to resample B_R times from the 120 time series, obtaining new factors for each draw, and re-estimating the VAR coefficients and the coefficients relating external regressors to the variables included in the VAR. For each draw, the reduced-form IRFs can be constructed at each horizon for any variable of interest. From these draws, a covariance matrix can be estimated. While BBE apply the bootstrap-after-bootstrap of Kilian (1998), the method receives virtually no discussion in their paper. Kilian's approach is conceived to avoid bias in the OLS estimates, and provides no explicit guidance on dealing with the complicated dynamics that must be replicated in re-drawing the factors. Given the necessity in that method of resampling based on estimated parameter values, it is unclear how to proceed without imposing some parametric restrictions. Further, since bias-correction is not a regular feature of applied work, the approach of Gospodinov & Ng (2013) is favored here.

Inference on H follows the approach of Müller (2013), who offers an analog for QML inference in the posterior sampling context. If the model is misspecified, posterior sampling will be from a likelihood with a different shape to the true density. In this case, a sandwich estimator can be employed, which weakly improves both frequentist and Bayesian risk. Müller's Σ_M/T is estimated with the sampling covariance from the MCMC. The score is estimated based on 1000 draws of $\sigma^2_{1:T}$ obtained conditional on the estimated parameters, as recommended for high-dimensional parameter models in Section 6 of his paper. The long-

run covariance of the score is then estimated using the equal-weighted-periodogram HAC estimator proposed by Lazarus, Lewis, & Stock (2017), with 8 degrees of freedom.

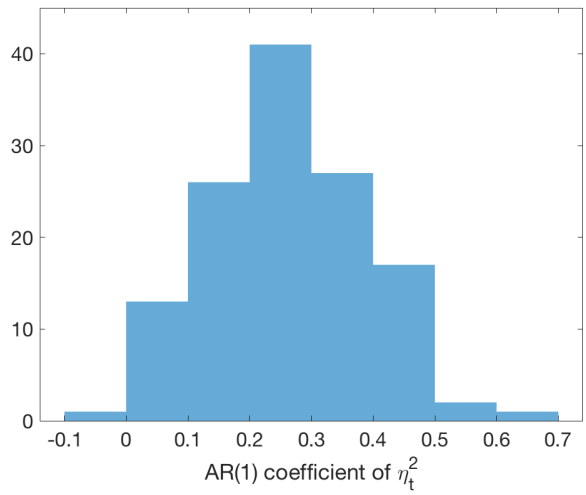
Combining the resulting structural covariance with that of the reduced-form IRFs completes a block diagonal covariance matrix. The delta method is then used to obtain the variance of a structural IRF. A similar approach can be used with an estimated covariance of the structural parameters corresponding to any of the other estimation strategies proposed in this paper.

B.5 GARCH parameters

In order to implement the GARCH filtered hybrid estimation approach, it is necessary to have a set of standard GARCH parameters appropriate to the setting at hand. Note that these values will likely depend on the type of variables considered and in particular the frequency of the observations. As with any persistent process, more frequent observations will exhibit much stronger autocovariance than more distant ones. For this paper, values were calculated based on the 120 monthly macro variables from 1960 to 2000 used in the FAVAR estimation of the empirical application. As the literature applying GARCH in these settings tends to model each structural variance as an independent GARCH(1,1) process, with no relation across variables, estimation occurred separately for each of the 120 series. While I am interested in fitting a GARCH(1,1) to structural shocks, these are not observed. Instead, I calculate the residuals from an AR(13) process (the same number of lags as in the empirical application, for consistency). Since these are simply linear combinations of underlying structural shocks, it is assumed that the GARCH parameters estimated will be representative of any GARCH dynamics displayed in innovations to macro variables more generally. Table 12 displays the mean parameter values obtained across the series fitting into each of the sub-groups discussed in BBE – first overall estimates, then slow and fast series, then each economic category. Note that for nine of the series, the model could not successfully be fitted due to violations of parameter constraints. The sub-division of estimates provided can be helpful to economists who are working with data corresponding to a particular category. For the purposes of this paper, in particular the empirical application, the overall estimates are the most relevant, as analysis proceeds on economy-wide factors. In addition, Figure 19 displays the distribution of estimated parameters for the overall dataset.

C Tables & Figures

Figure 1: Distribution of AR(1) coefficients of η_t^2



The η_t obtained as reduced-form innovations from AR(1) processes fitted to each of McCracken & Ng's 128 FRED-MD monthly time series. The figure displays the distribution of the implied AR(1) coefficients of η_t^2.

Table 1: The presence of off-diagonal elements

	$E\left[\varepsilon_{1t}\varepsilon_{2t} \mid t \in T\right]$	$E\left[\varepsilon_{1t}\varepsilon_{2t} \mid t \in A\right]$	$E\left[\varepsilon_{1t}\varepsilon_{2t} \mid t \in B\right]$
$H = I_2$	-0.001	-0.002	0.000
$H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	-0.001	0.414	-0.415

The table computes the conditional expectations noted via simulation. The variance matrix is I_2 for 500,000 observations and $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ for 500,000. The data is split into subsamples based on the trace of $\eta_t\eta_t'$. A is the subset of observations with trace above the median; B is A's complement.

Table 2: Data generating processes

	AR(1)	GARCH(1,1)	Markov
functional form	$\log\left(\sigma_t^2\right) = diag\left(\phi\right)\log\left(\sigma^2{}_{t-1}\right) + e_t$ e_t m.v. normal	$\sigma_{it}^2 = \mu_i\left(1-\psi_i-\Upsilon_i\right)$ $+\phi_i\sigma_{it-1}^2 + \Upsilon_i\varepsilon_{it-1}^2$	symmetric transition matrix
included ~~in study 1~~	✓		✓
empirical calibration	✓	✓	✓
"weak" calibration	✓	✓	
T	100, 200, 500, 1000	200	200
z_t distribution	$N\left(0,I\right)$, t_1	$N\left(0,I\right)$	$N\left(0,I\right)$
single dimension	✓		

The table describes the DGPs and various features considered across the simulation studies. "Study 1" refers to the study comparing tuning parameter choices for Rigobon-style estimators.

Table 3: Calibration of volatility processes

	AR(1)		
	ϕ	$E\left[e_te_t'\right]$	
empirical calibration	$(0.66, 0.91)'$	$\begin{bmatrix} 0.4 & 0.002 \\ 0.002 & 0.15 \end{bmatrix}$	
"weak" calibration	$(0.66, 0.91)'$	$\begin{bmatrix} 0.004 & 0.00002 \\ 0.00002 & 0.0015 \end{bmatrix}$	
	Single Dimension AR(1)		
	σ_1^2 (fixed)	ϕ	$E\left[e_t^2\right]$
empirical calibration	1.4253	0.91	0.15
	GARCH(1,1)		
	μ	ψ	Υ
empirical calibration	$(1.5951, 0.8907)'$	$(0.3068, 0.7236)'$	$(0.24, 0.1318)'$
"weak" calibration	$(1.0802, 0.4893)'$	$(0.3068, 0.7236)'$	$(0.024, 0.01318)'$
	Markov Switching		
	P_{trans}	V_{low}	V_{high}
empirical calibration	$\begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$	$\begin{bmatrix} 0.77 & 0 \\ 0 & 0.72 \end{bmatrix}$	$\begin{bmatrix} 2.08 & 0 \\ 0 & 2.36 \end{bmatrix}$

Calibration of the volatility processes used in the simulation studies. Values are based on estimates from a version of the empirical application with only two macro factors (plus the FFR). The "weak" versions are scaled to offer substantively less identifying variation, as discussed in Section D.1.

Table 4: Estimators tested

Identification scheme	Estimator	Summary
Rigobon	Sub-sample	13-month moving ave. trace, mean + 1 s.d. threshold
	$T/2$ split	subsample split at $T/2$
	Blocks	Minimum distance using 8 consecutive blocks of fixed length
Sentana & Fiorentini	GARCH	Independent GARCH(1,1) processes, maximum likelihood
	Kernel	Smoothed instantaneous reduced-form covariances, $b_T = 12$
TVV-ID	Quasi-likelihood	AR(1) log SV with correlated innovations (MCMC)
	Bayes	AR(1) with correlated innovations and generic priors (MCMC)
	GMM	2-step GMM using mean and first autocovariance of $\eta_t \eta_t'$
	Hybrid	GARCH(1,1) with calibrated GARCH and ARCH parameters

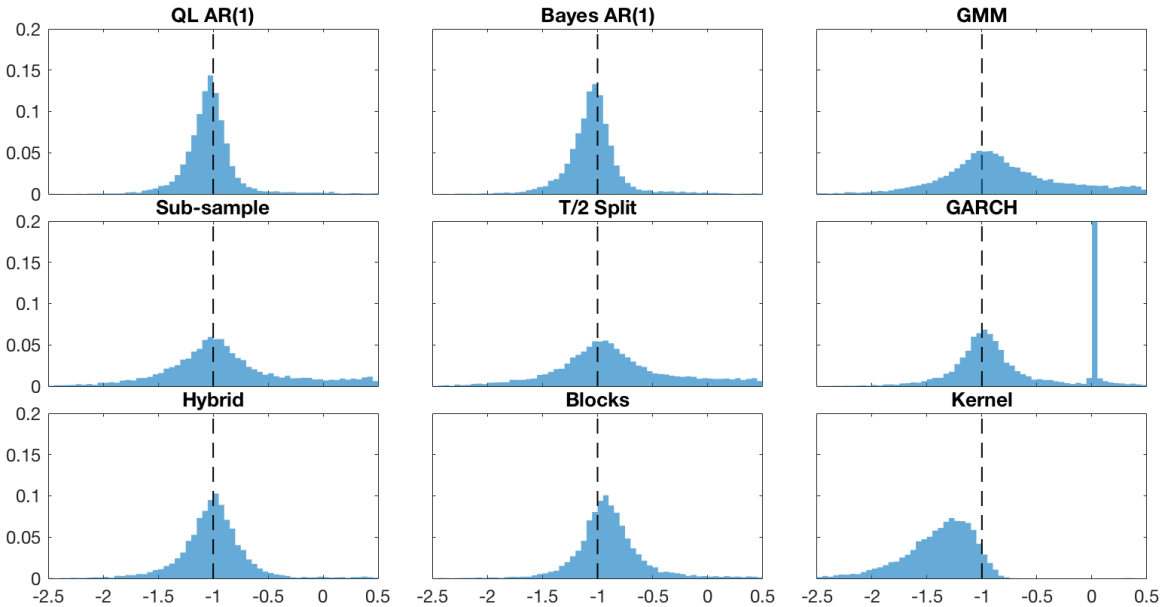
The table offers a brief description of estimators used under the three identification schemes considered. Note that some divisions between categories are not clear-cut. Additional details of DGPs are available in Section 5.2 and details of estimators in Appendix B.2.

Table 5: Median estimates for Markov-switching DGP

	window	1-period		7-period		13-period		oracle	
norm	threshold	H_{21}	H_{12}	H_{21}	H_{12}	H_{21}	H_{12}	H_{21}	H_{12}
trace	median	1.61	-1.72	1.01	-1.36	0.85	-1.25	0.50	-0.95
	$mean + 1\ s.d.$	1.46	-1.64	1.09	-1.42	0.91	-1.29		
$\bar{\eta}_1^2$	median	0.00**	-0.39**	0.05**	0.45**	0.09**	-0.50**		
	$mean + 1\ s.d.$	0.02**	-0.41**	0.04**	-0.43**	0.05**	-0.45**		

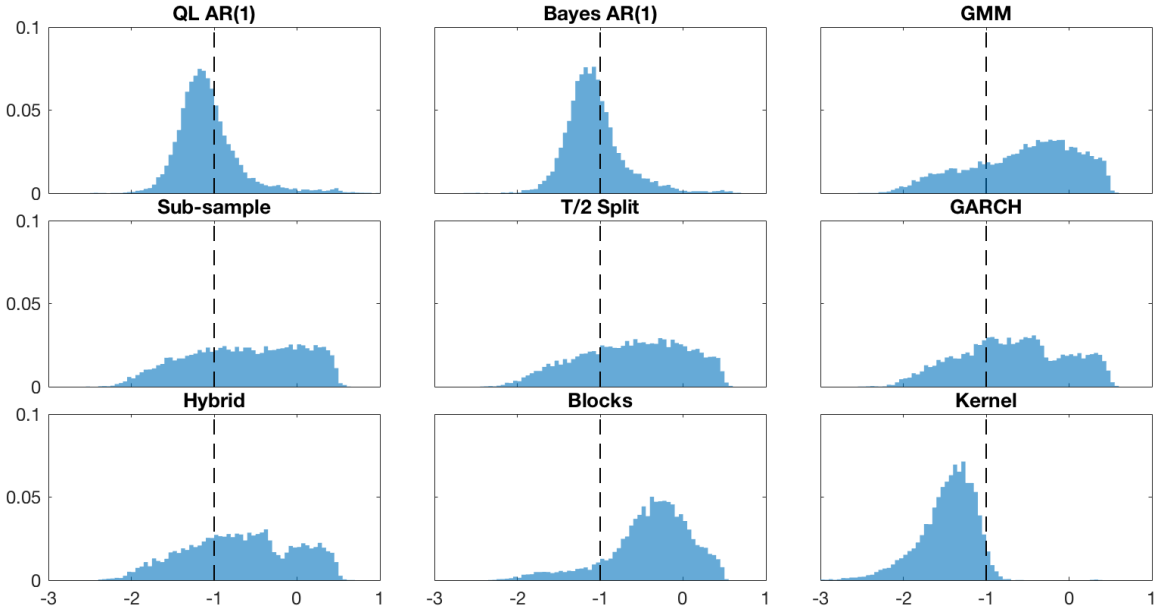
Median estimates of estimates for Rigobon-type estimators for the empirically-calibrated Markov-switching DGP, $T = 2000$, 10,000 draws. The window indicates the length of the rolling window over which variances were computed to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered "high variance". ** indicates the true value is not contained between in the 0.025 and 0.975 quantiles of the data, * indicates the same for 0.05 and 0.95. Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching H estimate to the true H to minimize L_2 norm.

Figure 2: Distribution of H_{12} estimates: empirically calibrated AR(1), $T = 200$



Distribution of estimates of H_{12} for various estimators for empirically-calibrated AR(1) DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. Note that the peak of the GARCH distribution is truncated to allow all estimators to be viewed on standardized axes.

Figure 3: Distribution of H_{12} estimates: empirically calibrated Markov Switching, $T = 200$



Distribution of estimates of H_{12} for various estimators for Markov switching DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Table 6: Median estimates and rejection rates

	QL AR(1)		Bayes AR(1)		GMM		Sub-sample (rolling)		Sub-sample ($T/2$)		GARCH		Hybrid		Blocking		Kernel	ratio (rolling)		ratio ($T/2$)		ratio (acv)	
	median	α	median	α	median	α	median	α	median	α	median	α	median	α	median	α	median	median	α	median	α	median	α
Single AR(1) volatility process																							
H_{21}	0.51	14.8	0.49	14.8	0.32	46.3	0.57	0.0	0.45	0.0	0.41	8.2	-0.43	9.0	-0.07	24.4	0.75	-	-	-	-	-	-
H_{12}	-0.93	16.6	-0.91	16.0	-0.91	43.5	-1.02	0.1	-0.89	0.0	0.57	5.8	0.46	7.6	-0.24	29.7	-1.28	-1.00	0.7	-0.90	0.0	-0.94	51.9
Empirically-calibrated AR(1) volatility process																							
H_{21}	0.60	14.0	0.61	15.5	0.44	27.9	0.51	0.0	0.47	0.0	0.43	30.4	0.55	12.6	0.49	9.7	1.06						
H_{12}	-1.04	8.3	-1.04	9.3	-0.89	27.6	-0.96	0.1	-0.92	0.0	-0.86	33.4	-1.00	10.5	-0.92	19.6	-1.34						
"Weak" AR(1) volatility process																							
H_{21}	0.50	11.2	0.50	8.0	-0.34	100.0	0.19	0.1	0.17	0.0	0.01	9.0	0.28	45.3	-0.01	68.0	1.10						
H_{12}	-0.95	9.1	-0.94	5.6	0.08	99.9	-0.59	0.1	-0.57	0.0	-0.36	6.9	-0.69	39.6	-0.33	68.5	-1.43						
Empirically-calibrated GARCH volatility process																							
H_{21}	0.70	32.6	0.68	32.3	0.47	26.4	0.39	0.2	0.41	0.0	0.54	6.7	0.54	12.4	0.40	30.3	0.85						
H_{12}	-1.13	15.8	-1.10	16.0	-0.86	28.2	-0.69	0.2	-0.75	0.0	-0.99	6.8	-0.99	12.6	-0.72	47.0	-1.40						
"Weak" GARCH volatility process																							
H_{21}	-0.09	63.0	-0.09	63.0	0.23	40.0	0.23	0.9	0.27	0.0	0.35	7.7	0.37	39.1	0.13	62.4	0.48						
H_{12}	0.21	67.1	0.22	67.0	-0.32	41.9	-0.31	0.8	-0.41	0.0	-0.60	9.2	-0.62	40.8	-0.12	76.6	-1.42						
Markov switching process																							
H_{21}	0.69	26.8	0.67	24.9	0.05	48.3	0.19	0.1	0.15	0.0	0.25	39.8	0.25	46.7	-0.02	58.4	1.13						
H_{12}	-1.13	24.4	-1.11	20.9	-0.45	42.2	-0.61	0.1	-0.58	0.0	-0.68	36.9	-0.68	39.9	-0.33	53.1	-1.42						

Median estimates for the full range of estimators for the specified DGPs. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. Rejection rates, α, are calculated by constructing 95% confidence intervals for each draw. For the MCMC methods, this is based on the covariance matrix calculated from chains. For GMM, the approach is the standard asymptotic variance estimator. For the subsample method and any ratio estimates, the approach is a block bootstrap with block length equal to 25. For GARCH, the Fisher information is used. For the hybrid, the QML variance is used. For the block estimator, a standard MD variance estimator is used.

Table 7: Median estimates and rejection rates

	QL AR(1)		Bayes AR(1)		GMM		Sub-sample (rolling)		Sub-sample ($T/2$)		GARCH		Hybrid		Blocking		Kernel
	median	α	median	α	median	α	median	α	median	α	median	α	median	α	median	α	median
Empirically-calibrated AR(1) volatility process, $T = 100$																	
H_{21}	0.56	11.6	0.55	13.7	0.36	34.3	0.44	0.1	0.44	0.0	0.32	24.4	0.53	20.2	0.45	14.4	1.06
H_{12}	-1.00	7.0	-0.99	7.7	-0.81	32.6	-0.88	0.1	-0.88	0.0	-0.73	25.8	-0.97	16.9	-0.86	23.0	-1.34
Empirically-calibrated AR(1) volatility process, $T = 500$																	
H_{21}	0.59	13.8	0.59	15.8	0.52	17.4	0.59	0.1	0.47	0.0	0.49	35.0	0.55	8.9	0.50	8.0	1.07
H_{12}	-1.03	9.4	-1.03	11.0	-0.97	18.3	-1.03	0.1	-0.92	0.0	-0.93	37.0	-1.00	7.3	-0.95	18.3	-1.34
Empirically-calibrated AR(1) volatility process, $T = 1000$																	
H_{21}	0.58	16.3	0.58	18.7	0.54	10.9	0.62	0.3	0.48	0.0	0.52	31.9	0.55	7.7	0.50	9.6	1.09
H_{12}	-1.02	10.6	-1.02	11.7	-0.99	12.2	-1.05	0.4	-0.93	0.0	-0.96	33.3	-1.00	6.4	-0.94	17.7	-1.34
Empirically-calibrated AR(1) volatility process, t_1-distributed shocks																	
H_{21}	0.55	41.7	0.55	18.7	-0.01	44.4	0.55	0.5	0.55	0.0	0.06	52.8	0.16	57.0	0.55	24.9	1.07
H_{12}	-0.99	39.1	-1.00	16.7	0.00	43.2	-1.00	0.24	-1.00	0.0	-0.03	52.9	0.00	59.9	-0.98	49.4	-1.09

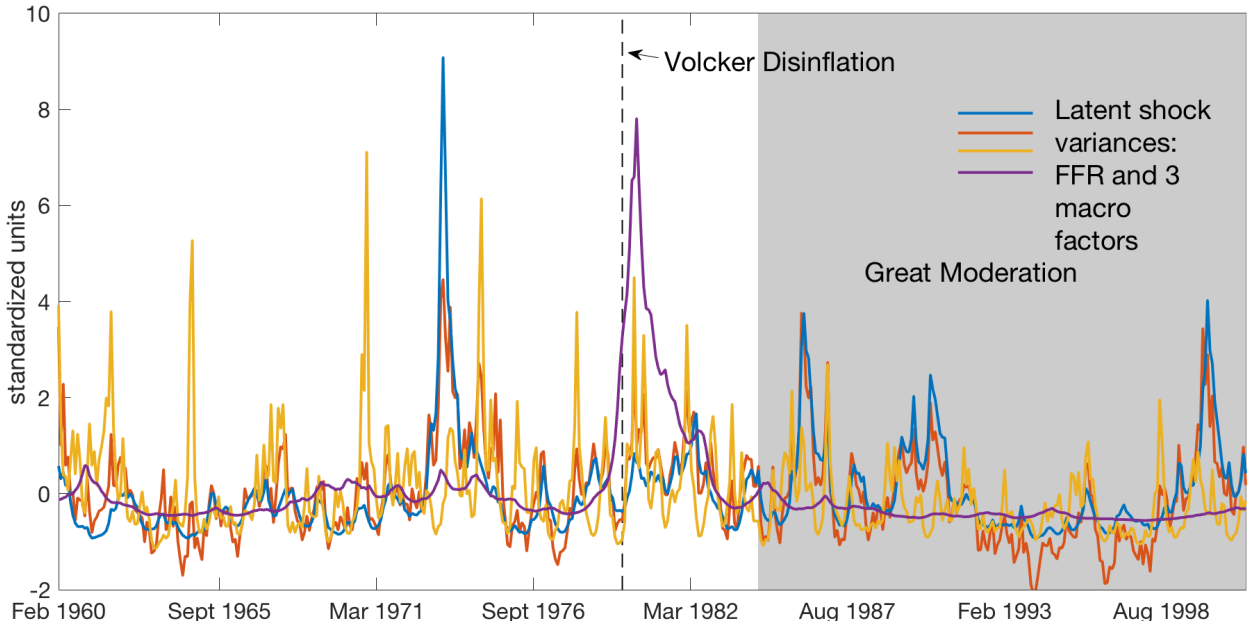
Median estimates for the full range of estimators for the specified DGPs. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. Rejection rates, α, are calculated by constructing 95% confidence intervals for each draw. For the MCMC methods, this is based on the covariance matrix calculated from chains. For GMM, the approach is the standard asymptotic variance estimator. For the subsample method, the approach is a block bootstrap with block length equal to 25. For GARCH, the Fisher information is used. For the hybrid, the QML variance is used. For the block estimator, a standard MD variance estimator is used.

Table 8: Candidates for $H^{(FFR)}$

	$H^{(FFR)}$			
factor 1	62.83	23.42	64.26	0.27
factor 2	22.70	-102.20	39.60	0.49
factor 3	3.38	10.20	53.68	-0.29
FFR	1	1	1	1

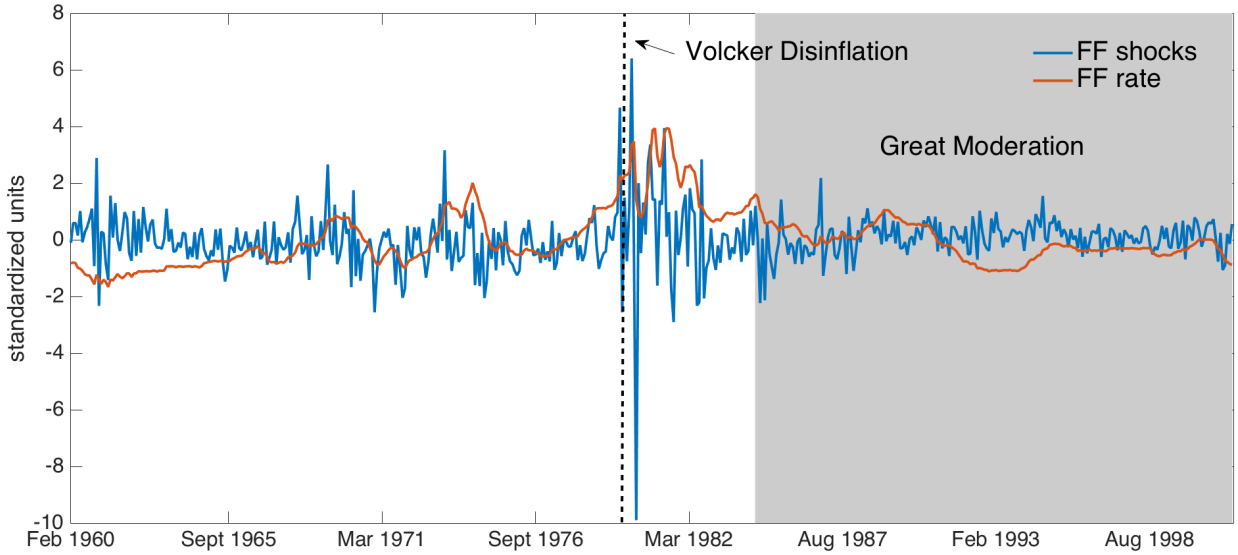
Candidates for the monetary policy column of H based on posterior medians quasi-likelihood AR(1) SV estimation via TVV-ID.

Figure 4: Variance paths



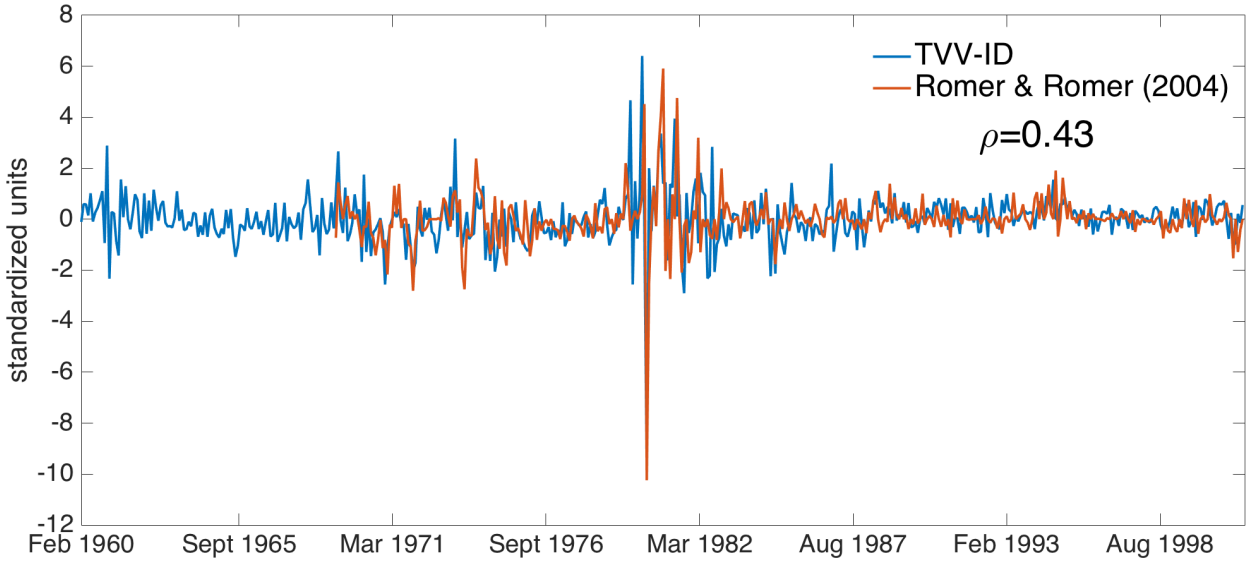
The lines plot the posterior medians for each of the four structural shock variances, standardized. Estimation via quasi-likelihood AR(1) approach. Volcker period and Great Moderation added for reference. The shock labeled as the monetary policy shock corresponds to the purple line. The colors correspond to those of the columns in Table 8.

Figure 5: Fed Funds and Fed Funds shocks



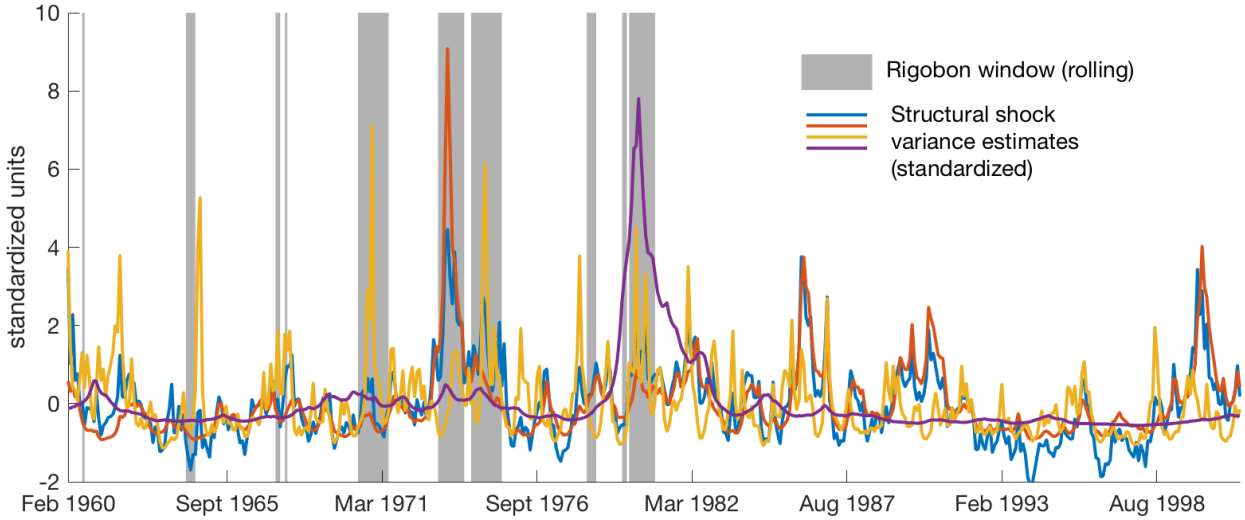
The monetary policy shock series is plotted (blue) with the the path of the Fed Funds rate (red). Volcker period and Great Moderation added for reference.

Figure 6: Shocks compared to Romer & Romer (2004)



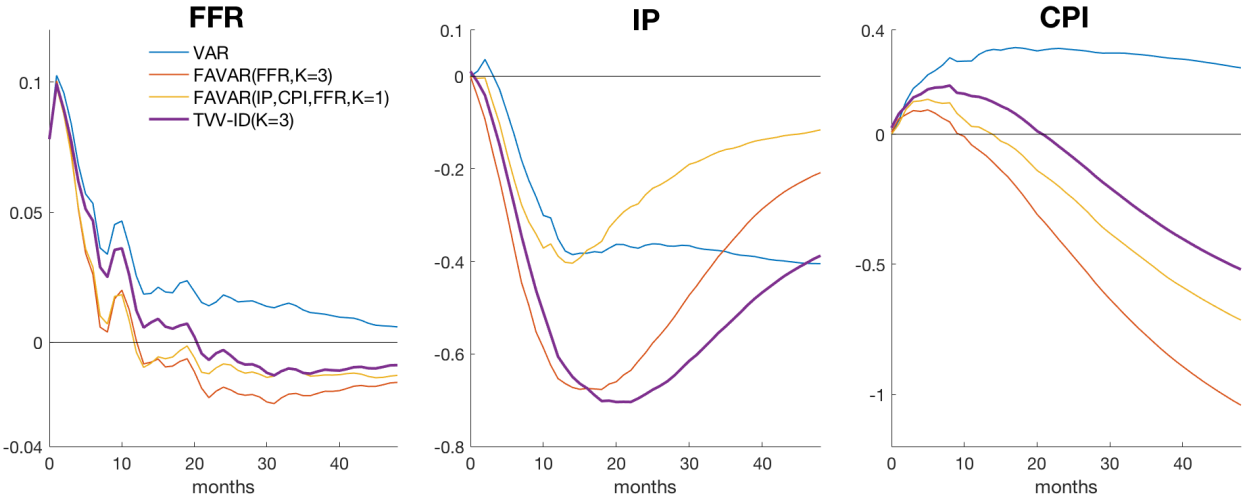
The blue line plots shocks estimated via TVV-ID using quasi-likelihood AR(1) SV. The red line plots the Romer & Romer (2004) shock series based on narrative data from FOMC minutes and the Greenbook. The correlation between the two series is 0.43, despite them being calculated from very different types of data.

Figure 7: Variance Paths with Rigobon Regimes



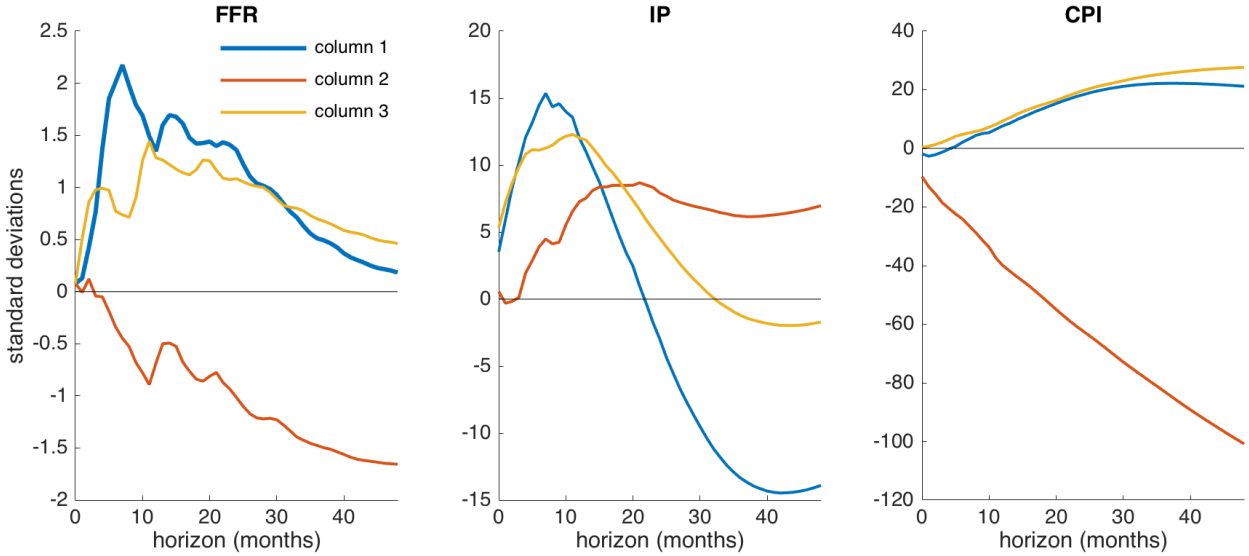
The lines plot the posterior medians for each of the four structural shock variances, as in Figure 4. The shaded areas correspond to the "high variance" regimes selected by a Rigobon rolling window selection rule based on a 13-month window and mean + 1 s.d. threshold for the trace.

Figure 8: Cholesky vs TVV-ID



Impulse responses to a 25 basis point shock to the Fed Funds rate, at a monthly frequency. The vertical axis is in standard deviations. TVV-ID quasi-likelihood AR(1) SV results in purple; additional lines replicated from BBE Figure I for comparison.

Figure 9: IRFs under alternative column labelings



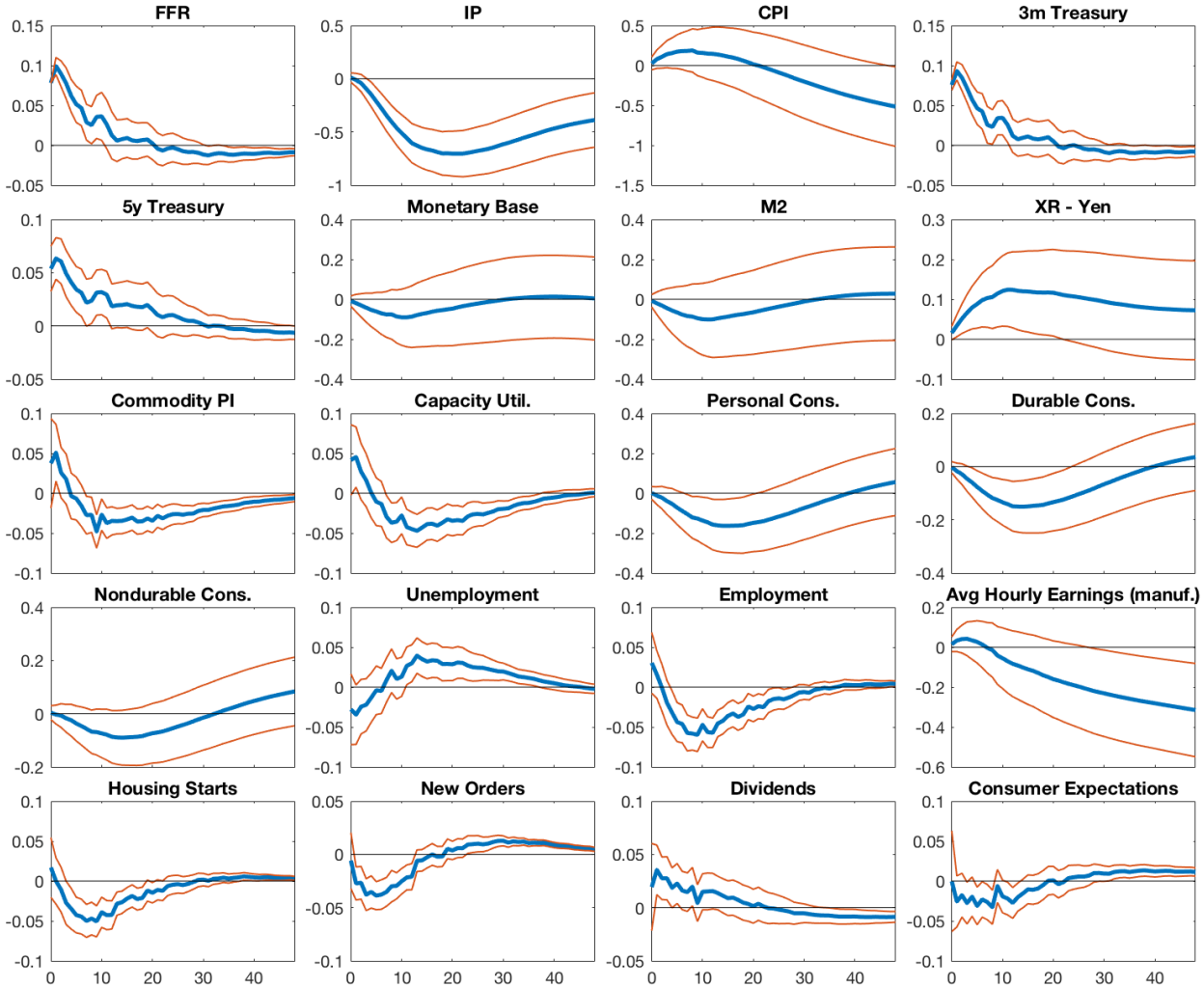
Impulse responses to a 25 basis point shock to the Fed Funds rate, at a monthly frequency. The vertical axis is in standard deviations. All lines are TVV-ID AR(1) quasi-likelihood results corresponding in color and order to the columns of Table 8.

Table 9: Estimates of $H^{(FFR)}$

	BBE (Cholesky)	TVV-ID
Factor 1	0	0.27
	–	(0.25)
Factor 2	0	0.50***
	–	(0.14)
Factor 3	0	-0.28**
	–	(0.13)
FFR	1	1
joint Wald test	–	75.26*** ($p < 0.001$)

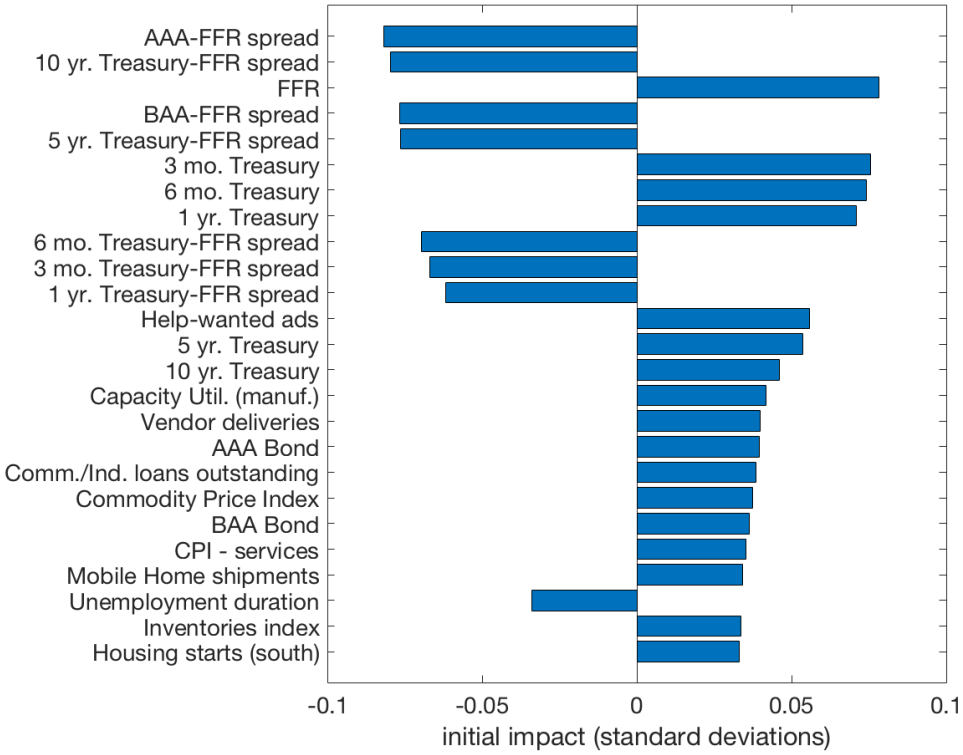
Comparison of the monetary policy column of H under Cholesky assumptions and TVV-ID posterior medians via quasi-likelihood AR(1) SV. Care must be taken in interpreting the scale of the factor coefficients as these are arbitrarily scaled factors, although they are standardized and estimated based on standardized macro time series. The joint Wald test is based on the Müller (2013) covariance estimator.

Figure 10: Impulse responses for important macro variables



Responses of 20 key macroeconomic variables to a 25 basis point Fed Funds shock identified via TVV-ID estimated by quasi-likelihood AR(1). The vertical axis is in standard deviations. Compare to Figure II of BBE. 95% confidence intervals are computed using the method described in Appendix B.4.

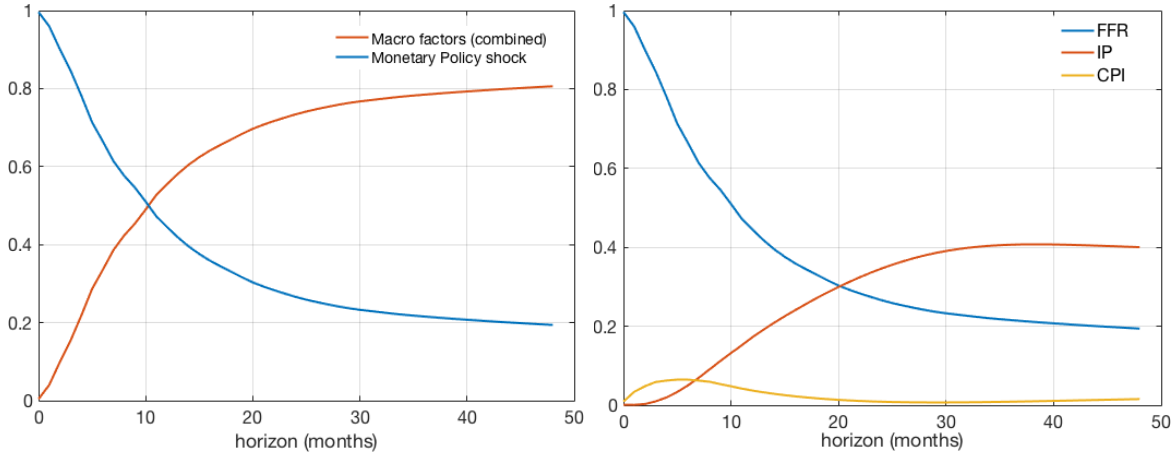
Figure 11: Initial Responses



The top 25 (by magnitude) initial responses from the 120-variable macro dataset to a 25 basis point monetary policy shock, in standard deviations. For comparison, the third row corresponds to the own-response of the FFR.

Figure 12: Forecast Error Variance Decompositions

(a) FEVD of monetary policy series (b) FEVD of macroeconomic series



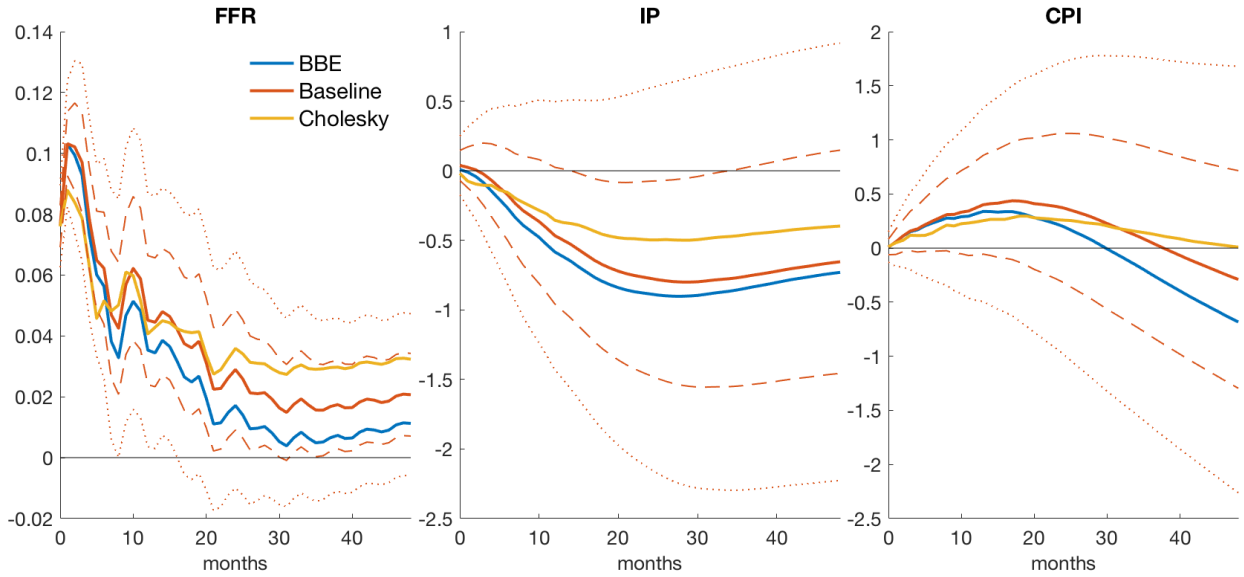
In panel 1, the blue line represents the forecast error variance decomposition of the FFR due to monetary policy shocks from 0-48 months; red plots the decomposition of the FFR due to the combined three macro factors. The FEVD is computed using the formula in Stock & Watson (2016). In panel 2, the blue line again represents the forecast error variance decomposition of the FFR due to monetary policy shocks; red decomposes IP due to monetary policy shocks, and yellow CPI.

Table 10: Estimates for Extensions

	Extended Sample	Two-dimensional policy		Stability of H	
	H^{MP}	H^{FF}	H^{G}	H^{MP}, '60-'84	H^{MP}, '84-'01
factor 1	0.17	0.43	0.73	0.27	0.44
factor 2	-0.41	0.33	0.73	0.67	0.16
factor 3	-0.09	0.16	0.37	-0.28	0.28
FFR	1	1	-0.04	1	1
5-year spread	-	0.29	1	-	-
marginal R^2	0.99	0.99	0.88	0.98	0.97
joint Wald test	131.33^{***} ($p < 0.001$)	29.44^{***} ($p < 0.001$)	46.79^{***} ($p < 0.001$)	124.25^{***} ($p < 0.001$)	0.01

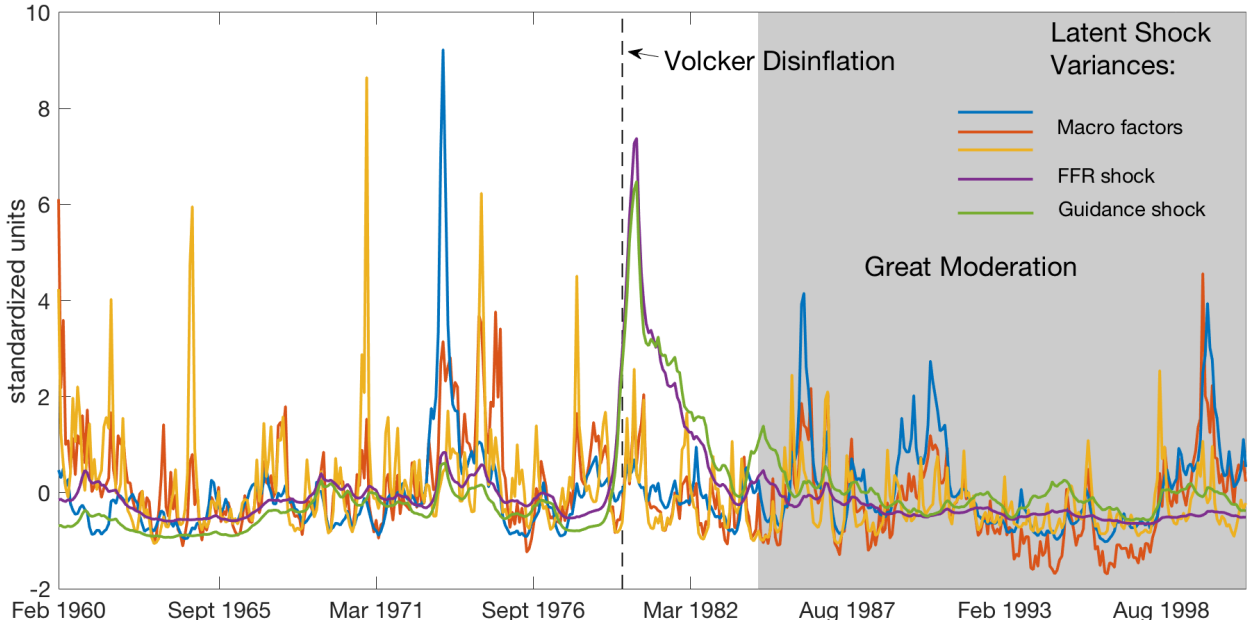
The Extended Sample column presents estimates for the monetary policy column of H obtained using TVV-ID on the McCracken & Ng FRED-MD data from 1960-2008. These compare to the Cholesky assumption of $(0, 0, 0, 1)'$. The marginal R^2 is reported that led to labeling selection of the monetary policy column. The test statistic and $p-$value for the joint Wald test of the Cholesky structure are reported. The variance of the estimator is computed using the Müller (2013) covariance estimator. The next two columns present estimates for Fed Funds shock and guidance shock columns of H for the 2-dimensional monetary policy model estimated on the BBE data. The details are as in the previous column. Note that the Cholesky null hypothesis for the Fed Funds columns is $(0, 0, 0, 1, 0)'$, and for the guidance column $(0, 0, 0, 0, 1)'$. The next two columns present the sub-sample estimates for the pre-1984 sample and post-1984 sample for the BBE data. The details are as before.

Figure 13: Extended sample IRFs



Impulse responses to a 25 basis point shock to the Fed Funds rate, at a monthly frequency, based on the McCracken & Ng 1960-2008 data. The vertical axis is in standard deviations. TVV-ID quasi-likelihood AR(1) SV results in red, with one standard deviation and 95% confidence intervals (dashed and dotted, respectively). The blue lines are IRFs using the BBE approach of only including slow-moving variation in the FAVAR and a Cholesky decomposition. The yellow lines are IRFs using a Cholesky decomposition but not restricting the factors in the FAVAR.

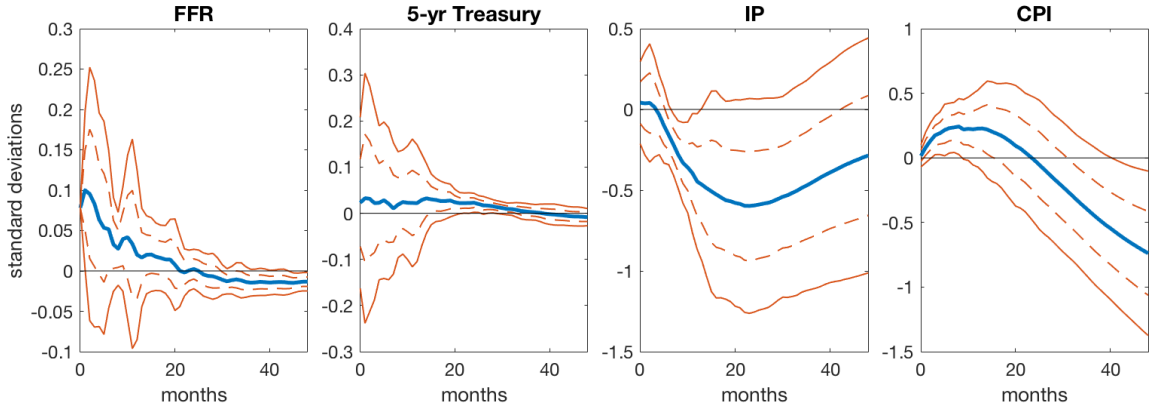
Figure 14: Variance paths for FFR and guidance shocks



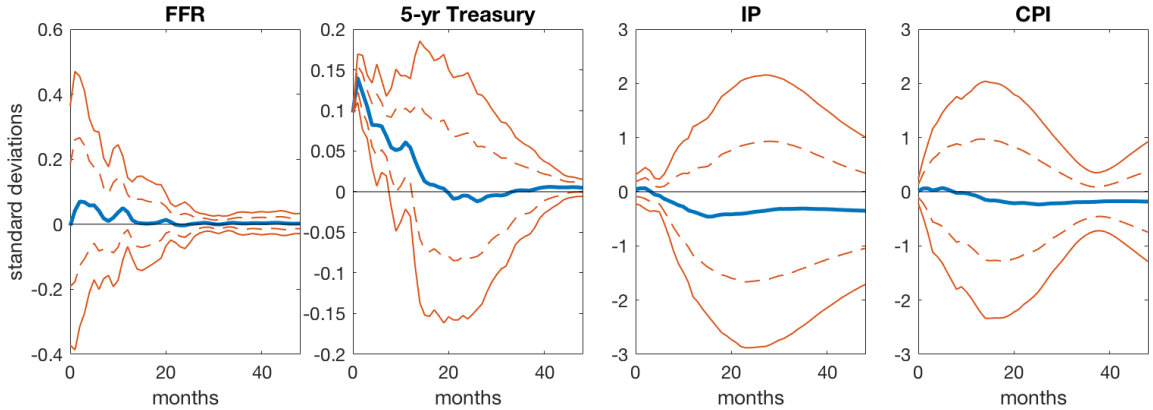
The lines plot the posterior medians for each of the five structural shock variances, standardized. The FFR shock variance is in purple and the guidance shock variance is in green. Estimation via quasi-likelihood AR(1) approach. Volcker period and Great Moderation added for reference.

Figure 15: IRFs for FFR and guidance shocks

(a) Fed Funds shocks



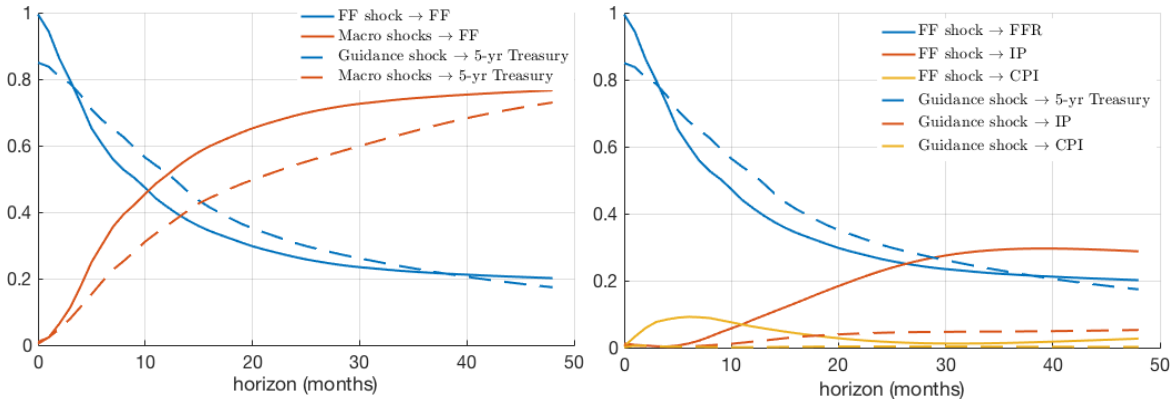
(b) Guidance shocks



Panel 1 plots IRFs to a 25 basis point FFR shock with one standard deviation and 95% confidence intervals, identified using TVV-ID and esttimated using AR(1) SV quasi-likelihood. The vertical axis is in standard deviations. Panel 2 plots IRFs to a 25 basis point guidance shock.

Figure 16: FEVDs for two-dimensional monetary policy

(a) FEVD of monetary policy series (b) FEVD of macroeconomic series



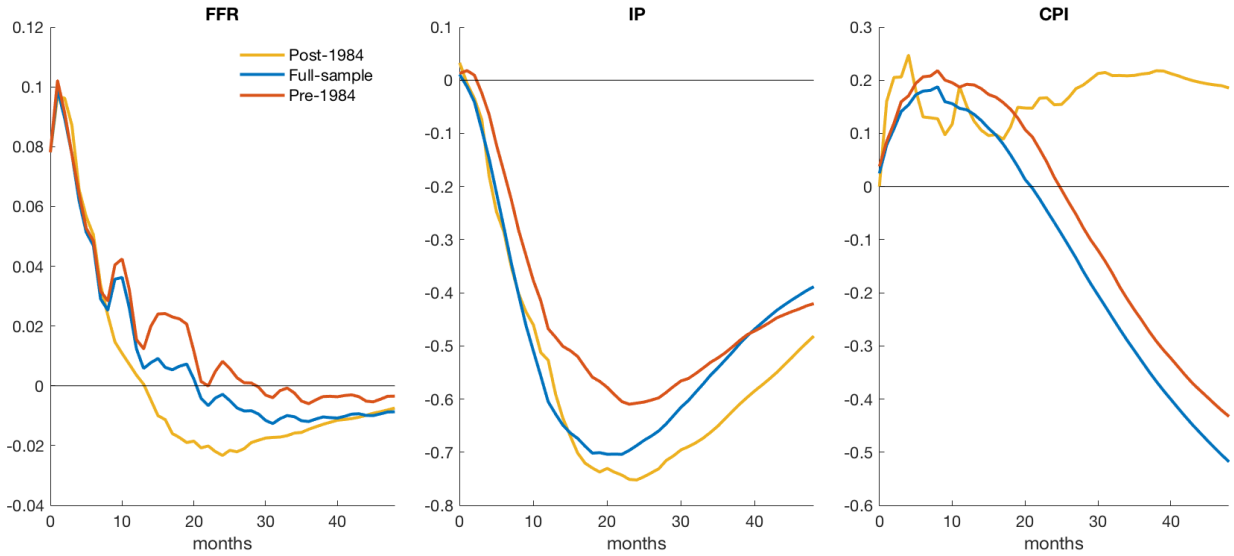
In panel 1, the blue lines plot FEVDs for the FFR and 5-yr treasury for variation due to their respective monetary policy shocks (FFR and guidance). The red lines plot the FEVDs for variation due to the combination of the three macroeconomic shocks. The FEVD is computed using the formula in Stock & Watson (2016). The solid lines in panel 2 plot the FEVD of the FFR, IP, and CPI due to the FFR shock, and the dotted lines plot the FEVD of the 5-yr Treasury, IP, and CPI to the guidance shock.

Table 11: Wald tests for the stability of H

	FIM	Müller
Full H	7.53 ($p = 0.82$)	6.87 ($p = 0.87$)
H^{MP}	1.10 ($p = 0.78$)	0.74 ($p = 0.86$)

Wald test statistics and $p-$values for the stability of H across the 1960-1984 and 1984-2001 sub-samples. Two null hypotheses are tested: that all elements of H are equal across sub-samples and that the monetary policy column of H is stable across sub-samples. Two covariance matrices are considered: an estimate of the efficient (under correct specification) FIM via MCMC sampling variation, and Müller's (2013) quasi-likelihood covariance estimator.

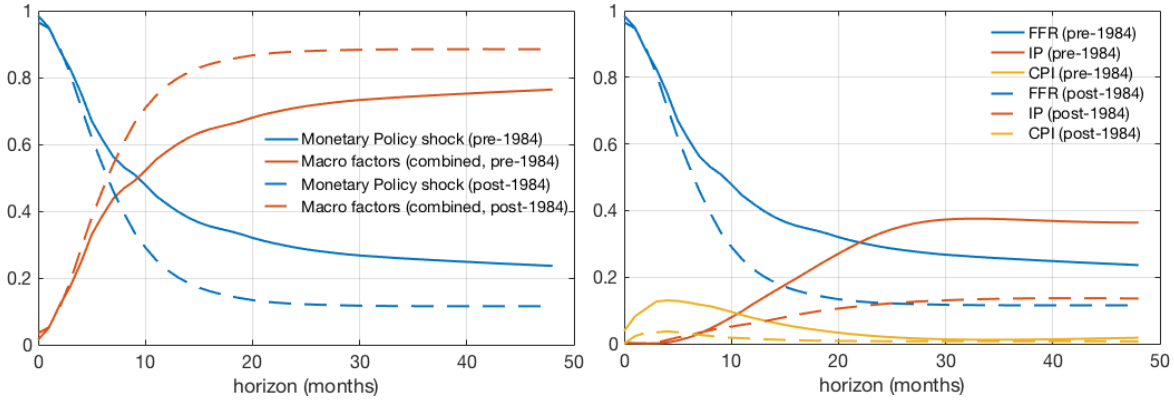
Figure 17: Sub-sample IRFs



The blue line plots the full-sample (1960-2001) IRFs for a 25 basis point monetary policy shock. The vertical axis is in standard deviations. The yellow line plots the responses estimated on the 1984-2001 sub-sample, allowing all reduced-form and structural parameters to change. The red line does the same for the 1960-1984 sub-sample.

Figure 18: Sub-sample FEVDs

(a) FEVD of monetary policy series (b) FEVD of macroeconomic series



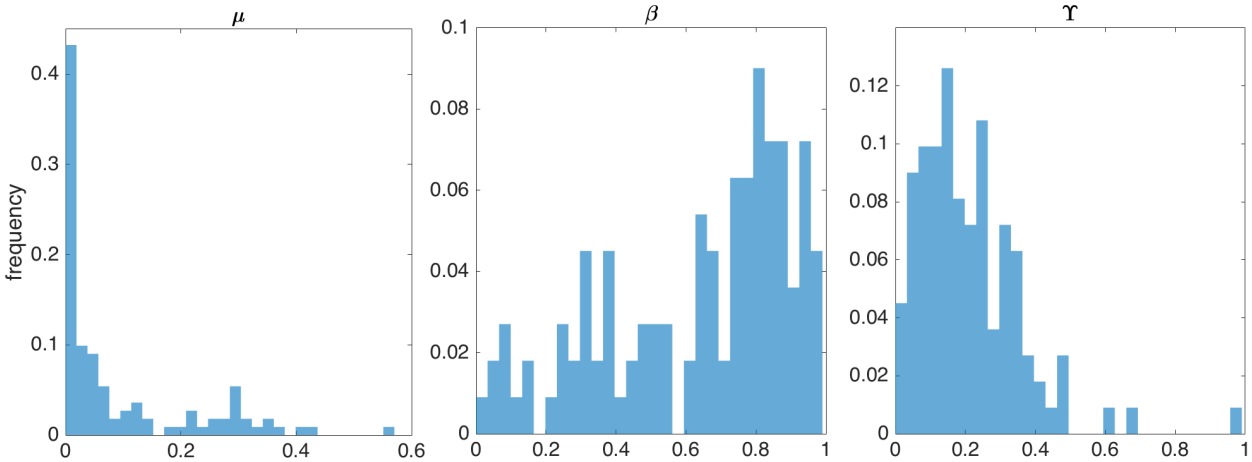
In panel 1, the blue lines represent the forecast error variance decomposition of the FFR due to monetary policy shocks for the 1960-1984 and 1984-2001 sub-samples; red plots the decomposition of the FFR due to the combined three macro factors. The FEVD is computed using the formula in Stock & Watson (2016). In panel 2, the blue lines again represents the forecast error variance decomposition of the FFR due to monetary policy shocks; red decomposes IP due to monetary policy shocks, and yellow CPI.

Table 12: Estimated GARCH(1,1) parameters

Category	n	μ	β	Υ
Overall	120	0.0919	0.6378	0.2101
Slow	70	0.1117	0.6244	0.2134
Fast	50	0.0658	0.6553	0.2056
Real Output and income	21	0.1992	0.5075	0.2168
Employment and hours	27	0.0559	0.6805	0.2126
Consumption	5	0.1252	0.6737	0.1827
Housing starts and sales	7	0.0491	0.4921	0.2168
Real inventories, orders, and unfilled orders	5	0.1209	0.5442	0.1523
Stock prices	7	0.1024	0.7858	0.1010
Exchange rates	4	0.0117	0.8463	0.1481
Interest rates	15	0.0025	0.7393	0.2479
Money and credit aggregates	10	0.1572	0.5157	0.2770
Price indices	16	0.0681	0.6965	0.1754
Average hourly earnings	2	0.3529	0.0774	0.5368
Expectations	1	0.0006	0.8690	0.1310

Mean GARCH(1,1) parameters calibrated from AR(13) innovations from 120 monthly macro time series. Categories follow those used in BBE. The time series used are the residuals from an AR(13) estimated variable-by-variable.

Figure 19: Distribution of GARCH(1,1) parameter estimates



Distribution of GARCH(1,1) parameters calibrated from AR(13) innovations 120 monthly macro time series. Categories follow those used in BBE. The time series used are the residuals from an AR(13) estimated variable-by-variable.

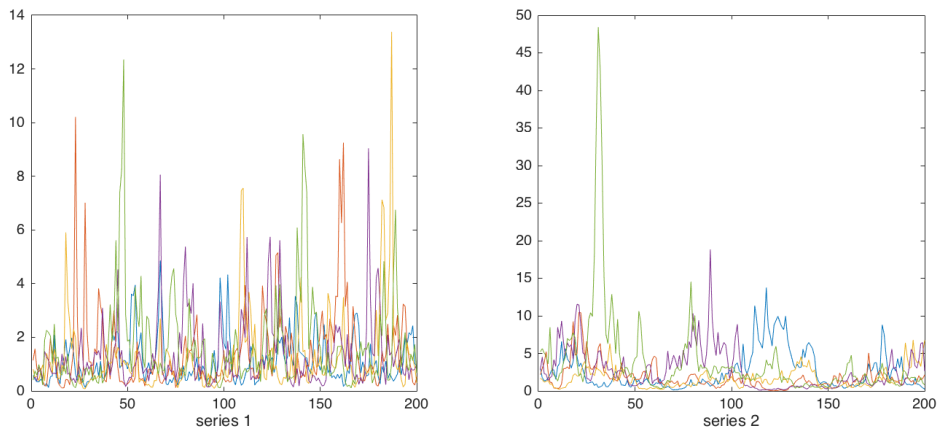
D Online Appendix

D.1 Paths for the calibration of the DGPs

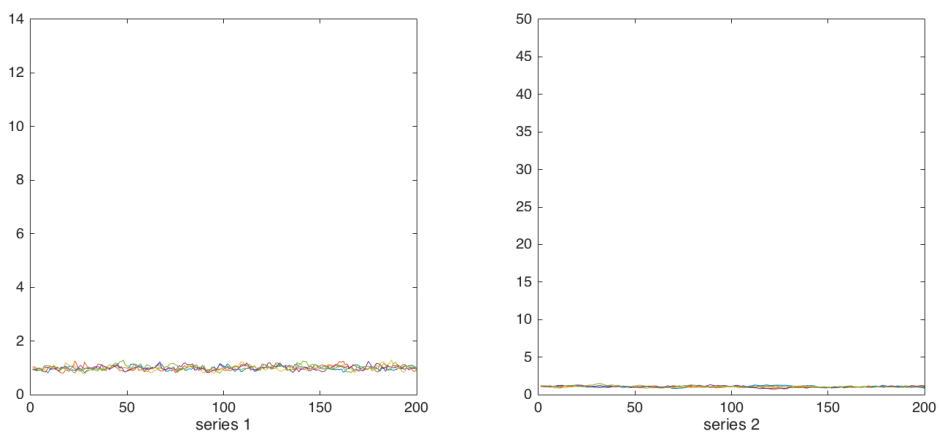
Figures 20 and 21 display sample paths for the variances for the empirically calibrated specifications and the "weak" specifications used to generate the simulation study, for the AR(1) SV and GARCH(1,1) processes respectively. Note that for each DGP, the "weak" paths show much smaller fluctuation about the mean, and that the difference in scale of fluctuation between empirical and weak is comparable across DGPs.

Figure 20: Calibration Paths, AR(1) log SV

(a) Empirical Calibration



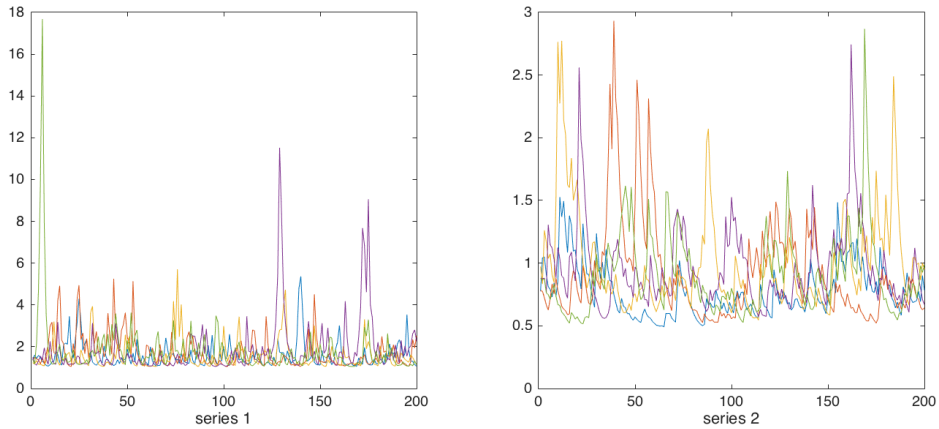
(b) "Weak" Calibration



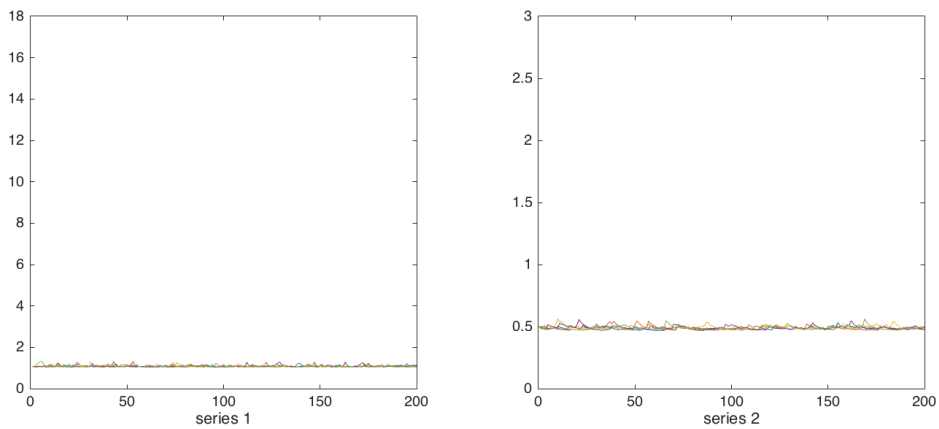
Comparison of sample variance paths for the log AR(1) SV process for empirical and "weak" calibrations. In the top panel, the paths are calibrated based on the first two factors in a reduced-dimension version of the empirical application. In the lower panel, the paths are for the weak calibration, which divides the variance innovation covariance matrix by 100.

Figure 21: Calibration Paths, GARCH(1,1)

(a) Empirical Calibration



(b) "Weak" Calibration



Comparison of sample variance paths for the GARCH(1,1) process for empirical and "weak" calibrations. In the top panel, the paths are calibrated based on the first two factors in a reduced-dimension version of the empirical application. In the lower panel, the paths are for the weak calibration, which divides the ARCH parameters by 10.

D.2 Additional simulation results

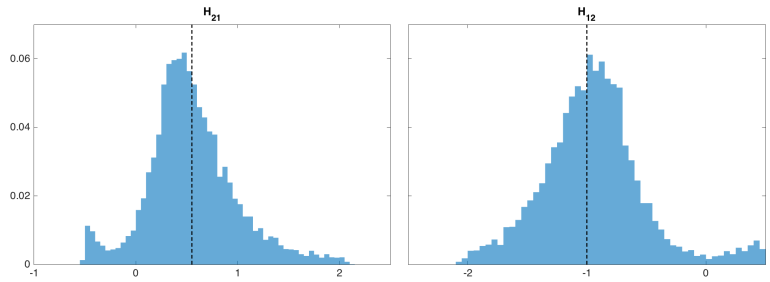
These tables and figures present additional simulation results discussed in the text. In particular, for Study 1, they include oracle histograms, the table for the AR(1) SV process, and histograms for all estimators for both H parameters for both DGPs. For Study 2, they include histograms for both estimates for all DGPs and estimators, expanding on the tables of summary statistics (Tables 6 & 7) in the main text and Figures 2 & 3.

Table 13: Median estimates for SV DGP

	window	1-period		7-period		13-period	
norm	threshold	H_{21}	H_{12}	H_{21}	H_{12}	H_{21}	H_{12}
trace	median	1.43	-1.60	0.79*	-1.20	0.68	-1.11
	$mean + 1\ s.d.$	1.04	-1.38	0.67	-1.11	0.62	-1.07
$\bar{\eta}_2^2$	median	0.03**	-0.42**	0.27**	-0.71**	0.36**	-0.81**
	$mean + 1\ s.d.$	0.17**	-0.61**	0.38**	-0.84**	0.43	-0.89

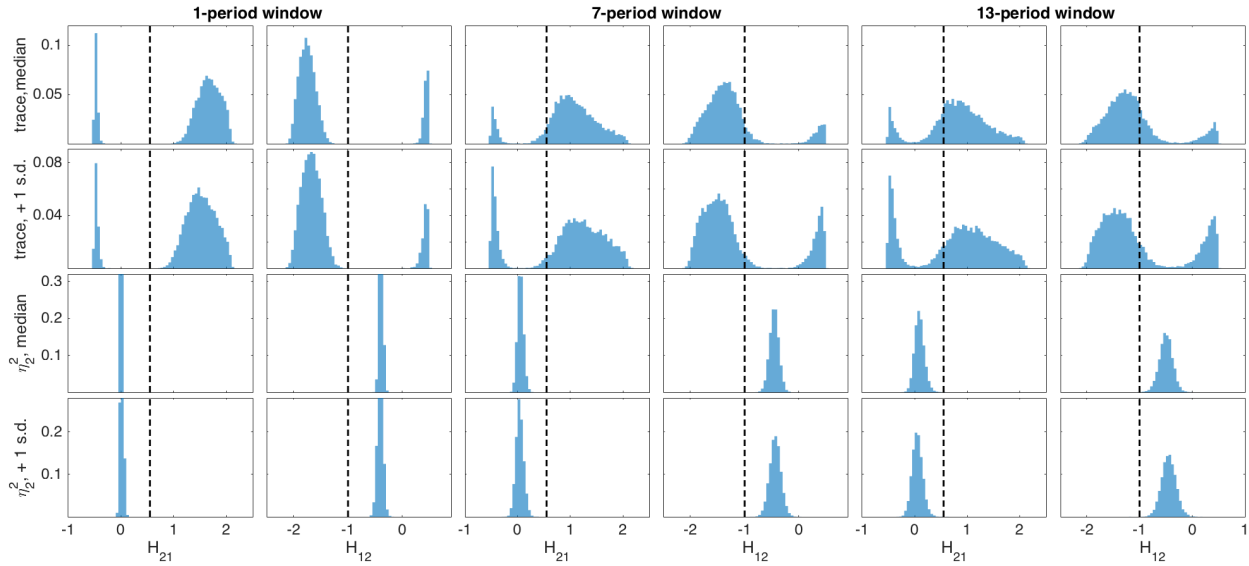
Median estimates of estimates for Rigobon-type estimators on the empirically-calibrated AR(1) SV DGP, $T = 2000$, 10,000 draws. The window indicates the length of the rolling window over which variances were compared to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered "high variance". ** indicates the true value is not contained between in the 0.025 and 0.975 quantiles of the data, * indicates the same for 0.05 and 0.95. Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 22: Oracle histograms for Markov-switching DGP



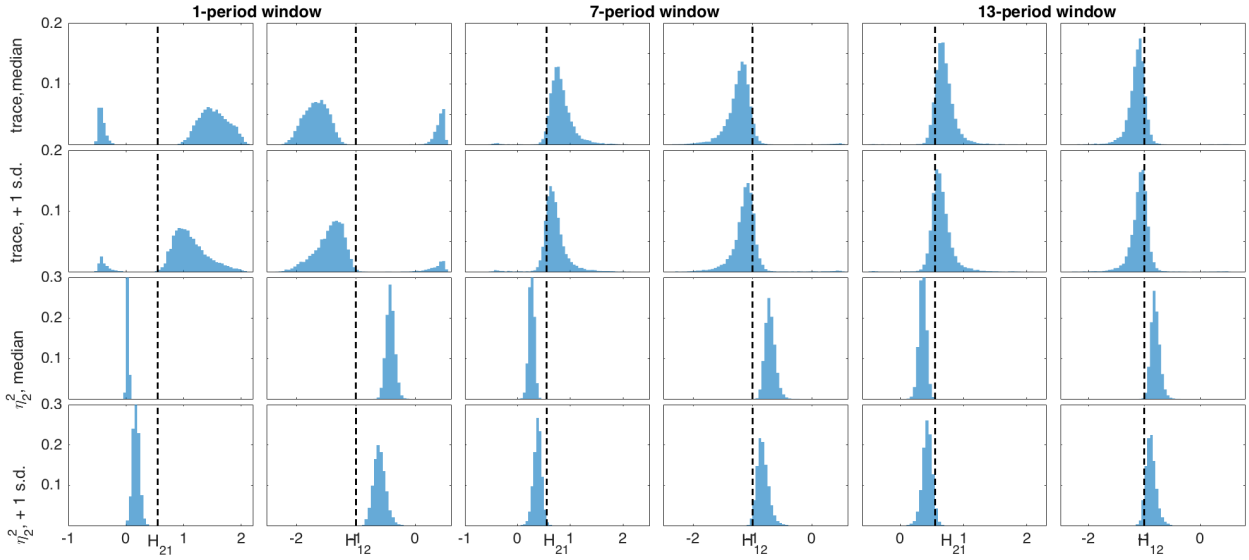
Distribution of estimates given knowledge of the true Markov switching dates, $T = 2000$, 10,000 draws. Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. The lack of normality is due to remaining imperfections in labeling accuracy.

Figure 23: Histograms for Markov-switching DGP



Distribution of estimates of estimates for Rigobon-type estimators on the empirically-calibrated Markov-switching DGP, $T = 2000$, 10,000 draws. The window indicates the length of the rolling window over which variances were computed to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered "high variance". Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

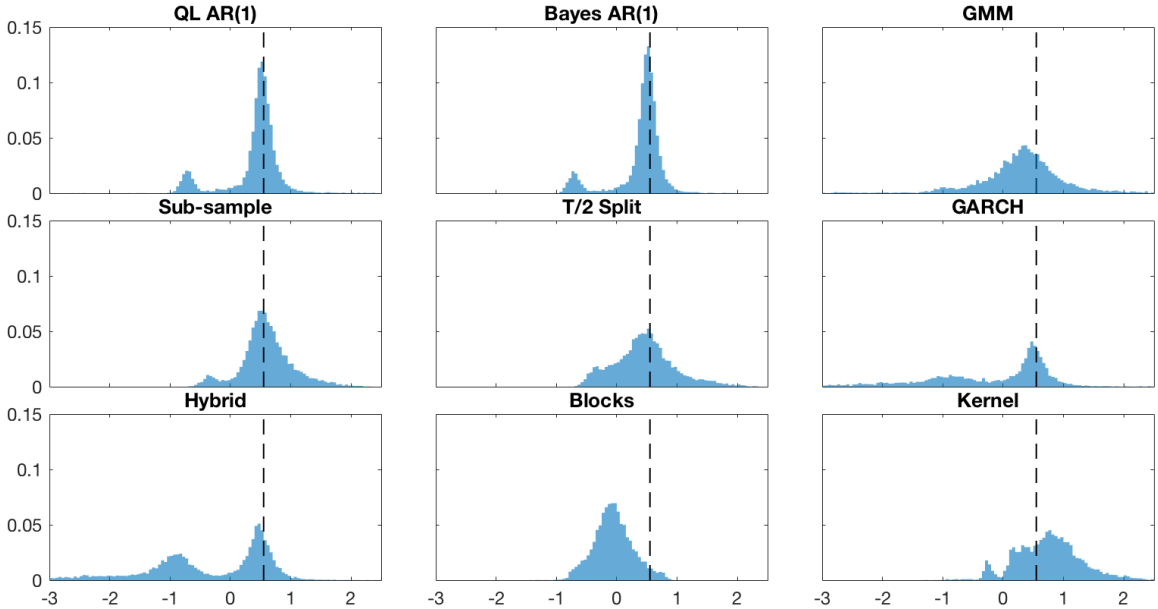
Figure 24: Histograms for SV DGP



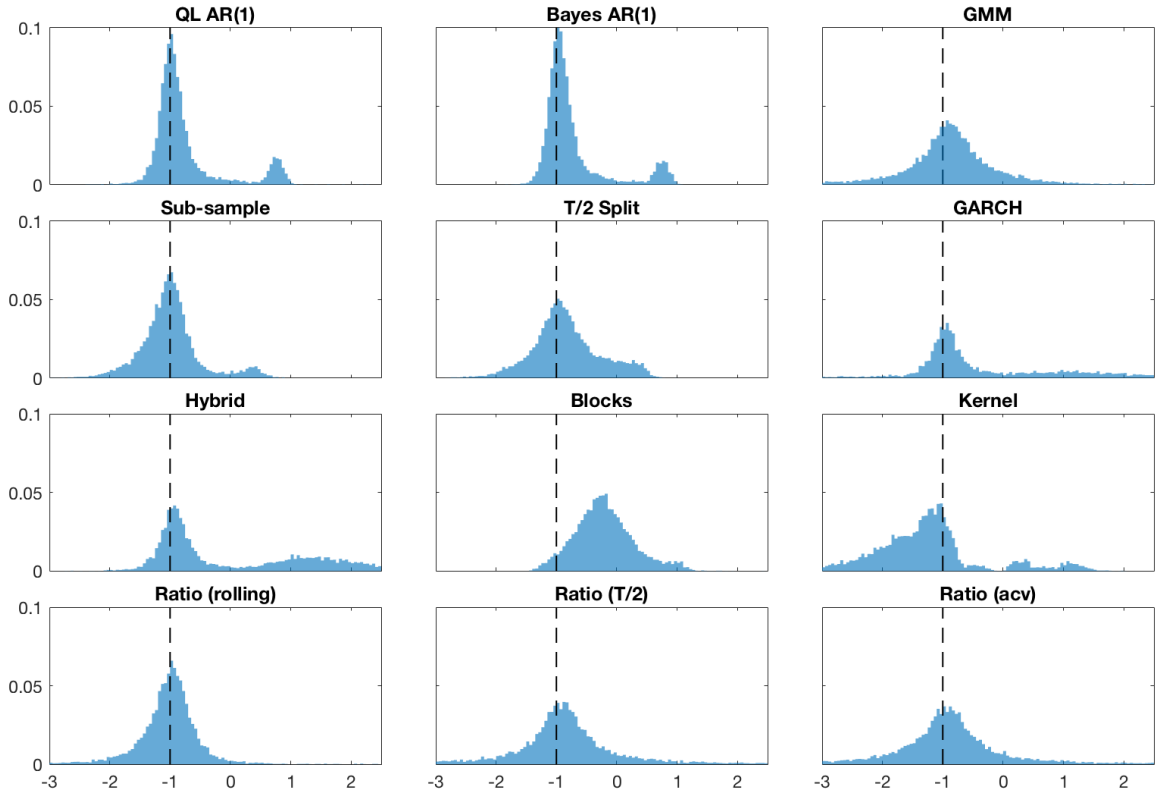
Distribution of estimates of estimates for Rigobon-type estimators on the empirically-calibrated AR(1) DGP, $T = 2000$, 10,000 draws. The window indicates the length of the rolling window over which variances were computed to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered "high variance". Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 25: Distribution of estimates: empirically calibrated single AR(1), $T = 200$

(a) Distribution of H_{21} estimates



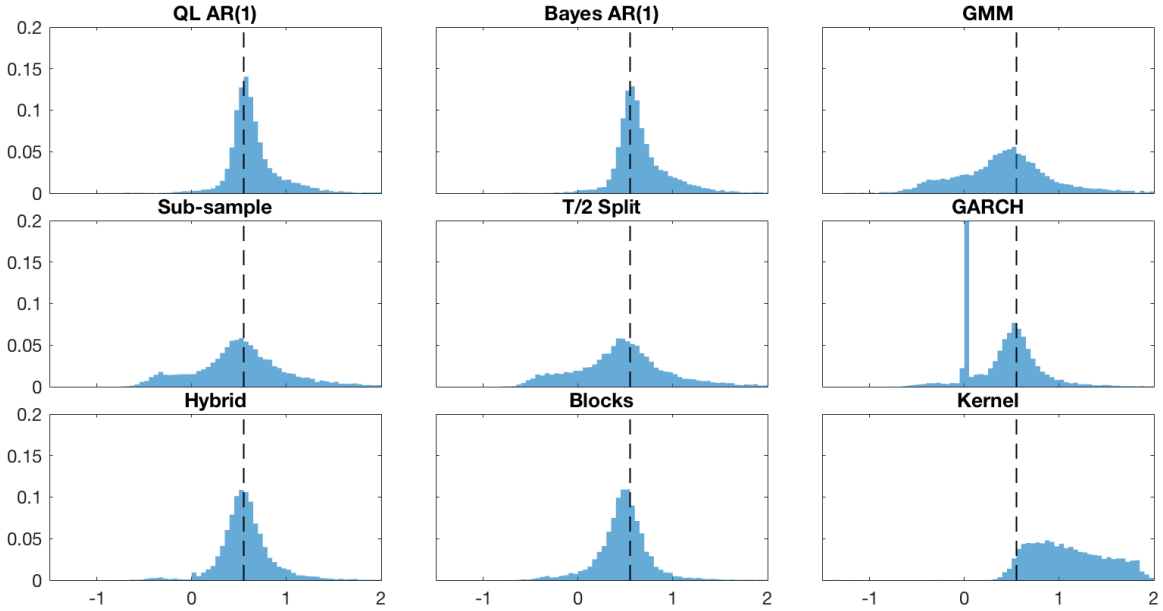
(b) Distribution of H_{12} estimates



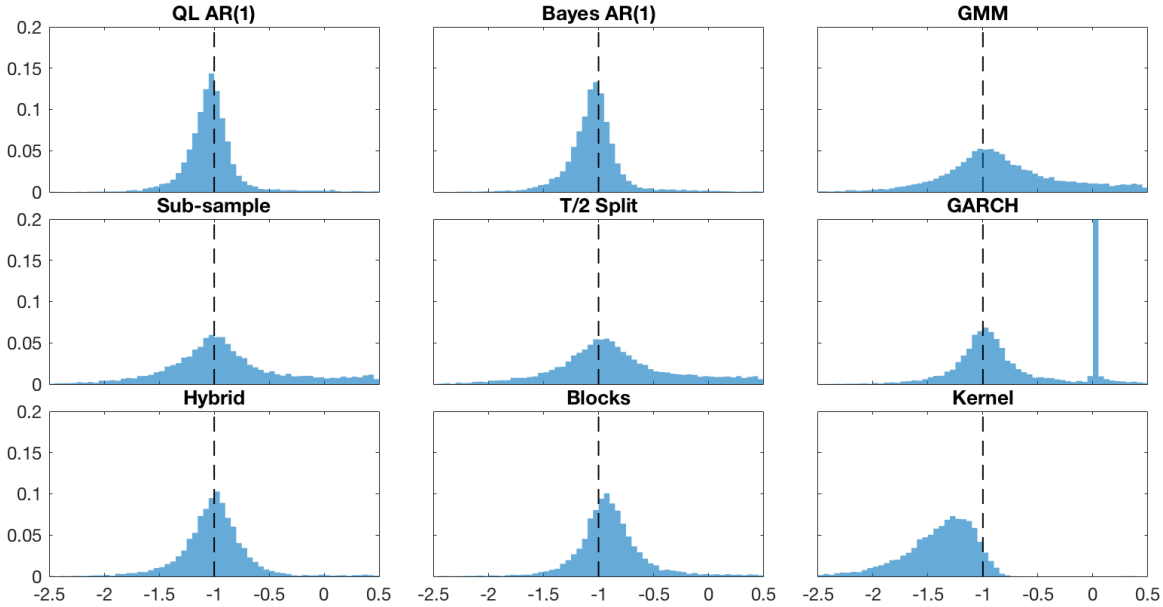
Distribution of estimates of H_{21} and H_{12} for various estimators for empirically-calibrated AR(1) SV DGP in a single dimension, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 26: Distribution of estimates: empirically calibrated AR(1), $T = 200$

(a) Distribution of H_{21} estimates



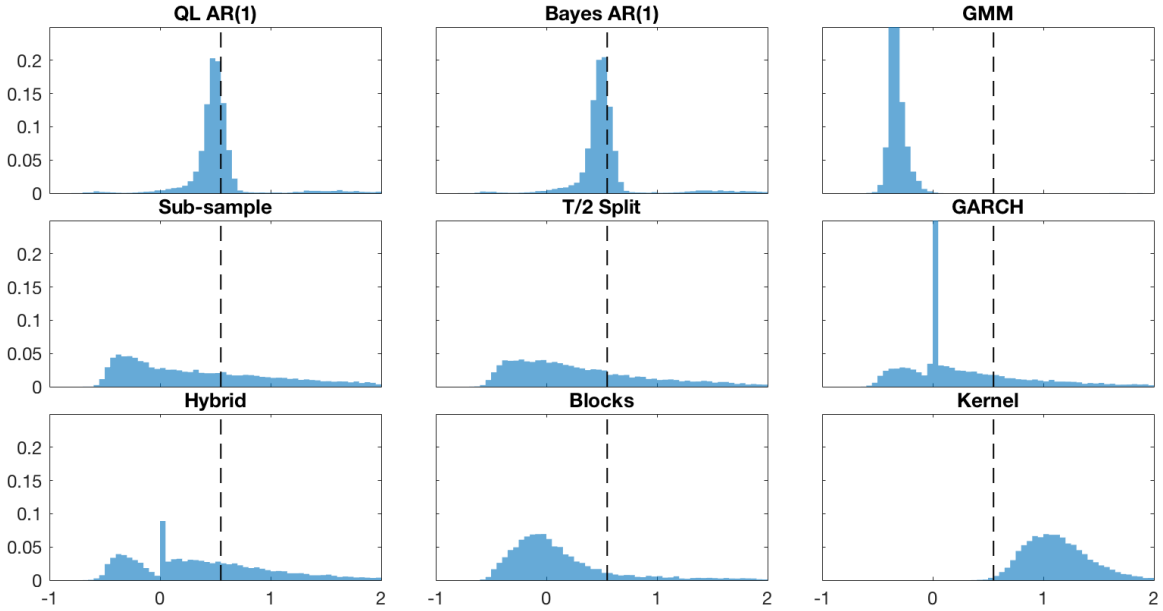
(b) Distribution of H_{12} estimates



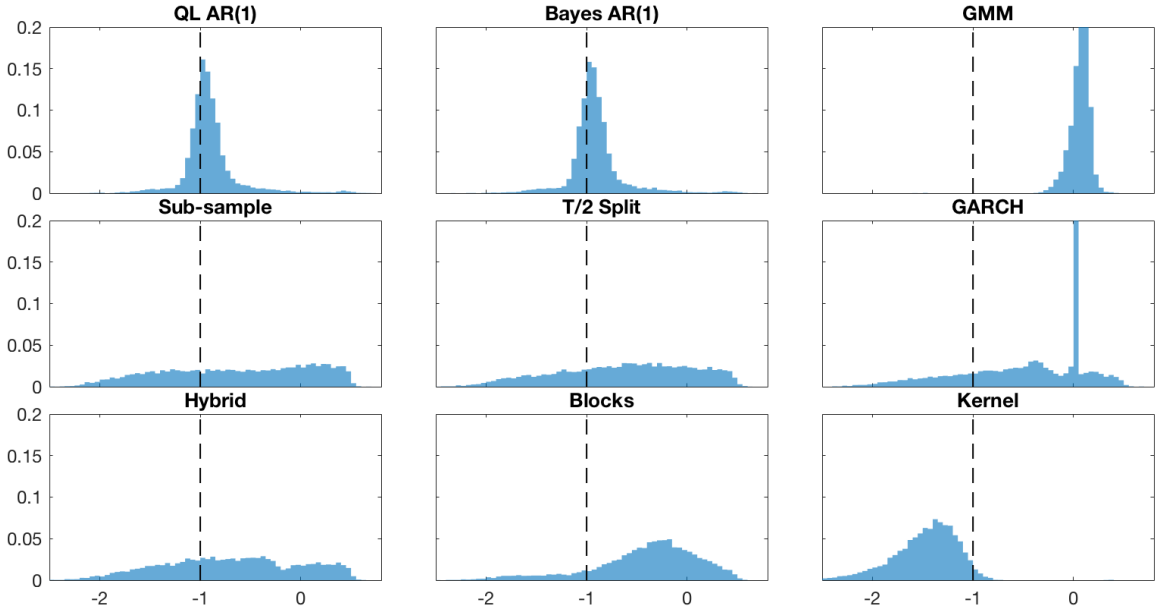
Distribution of estimates of H_{21} and H_{12} for various estimators for empirically-calibrated AR(1) SV DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. Note that the peak of the GARCH distribution is truncated to allow all estimators to be viewed on standardized axes.

Figure 27: Distribution of estimates: "weak" AR(1) SV, $T = 200$

(a) Distribution of H_{21} estimates



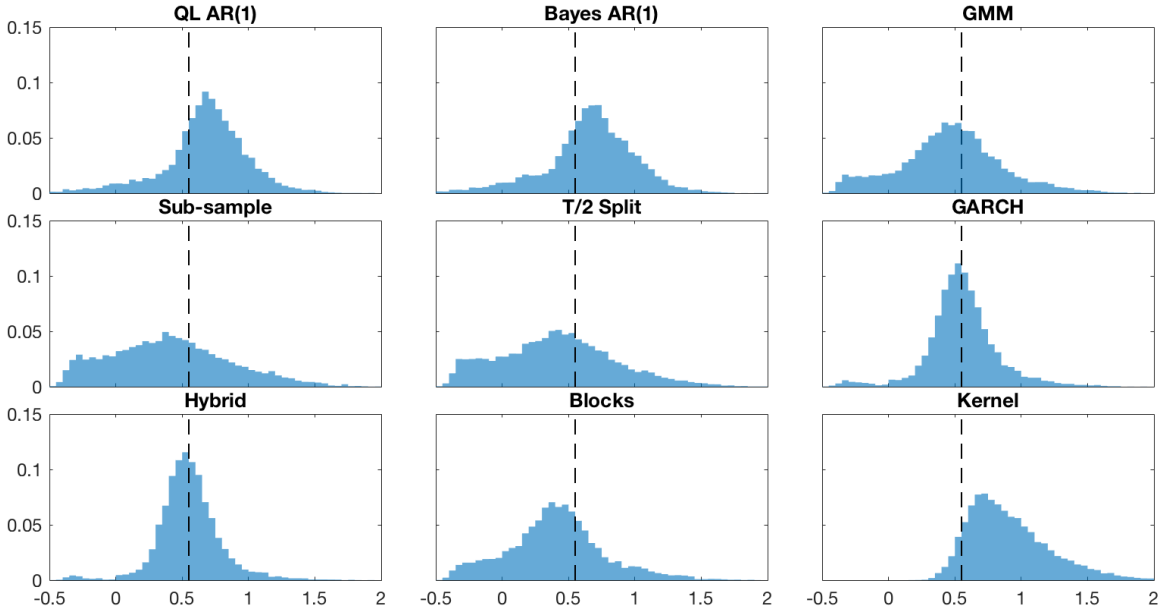
(b) Distribution of H_{12} estimates



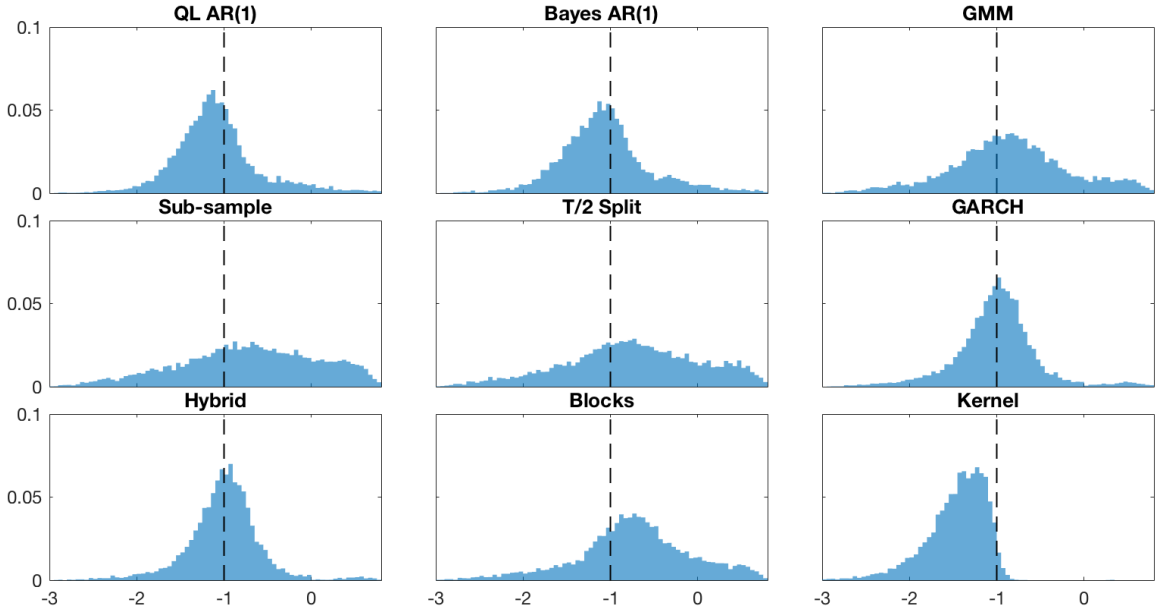
Distribution of estimates of H_{21} and H_{12} for various estimators for "weak" AR(1) SV DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. Note that the peak of the GARCH distribution is truncated to allow all estimators to be viewed on standardized axes.

Figure 28: Distribution of estimates: empirically calibrated GARCH, $T = 200$

(a) Distribution of H_{21} estimates



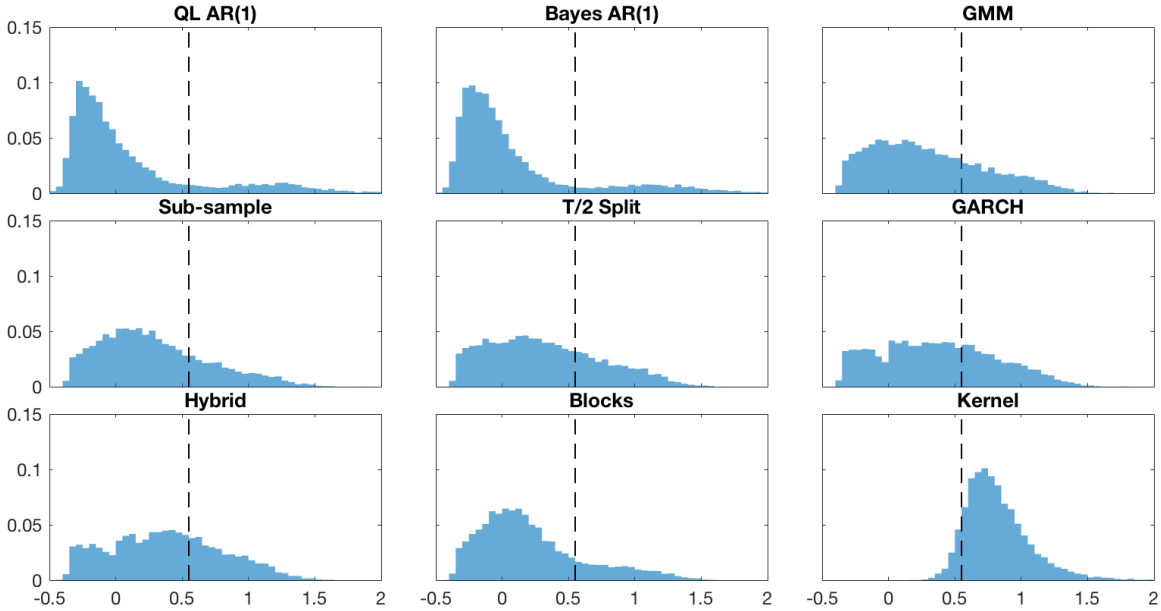
(b) Distribution of H_{12} estimates



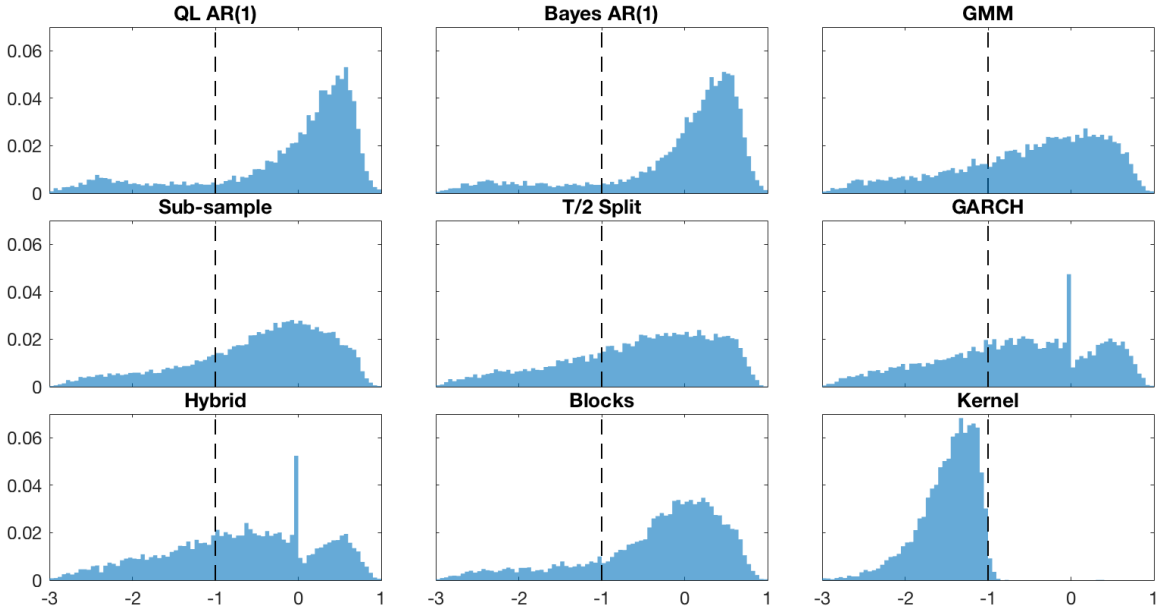
Distribution of estimates of H_{21} and H_{12} for various estimators for empirically-calibrated GARCH(1,1) DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 29: Distribution of estimates: "weak" GARCH, $T = 200$

(a) Distribution of H_{21} estimates



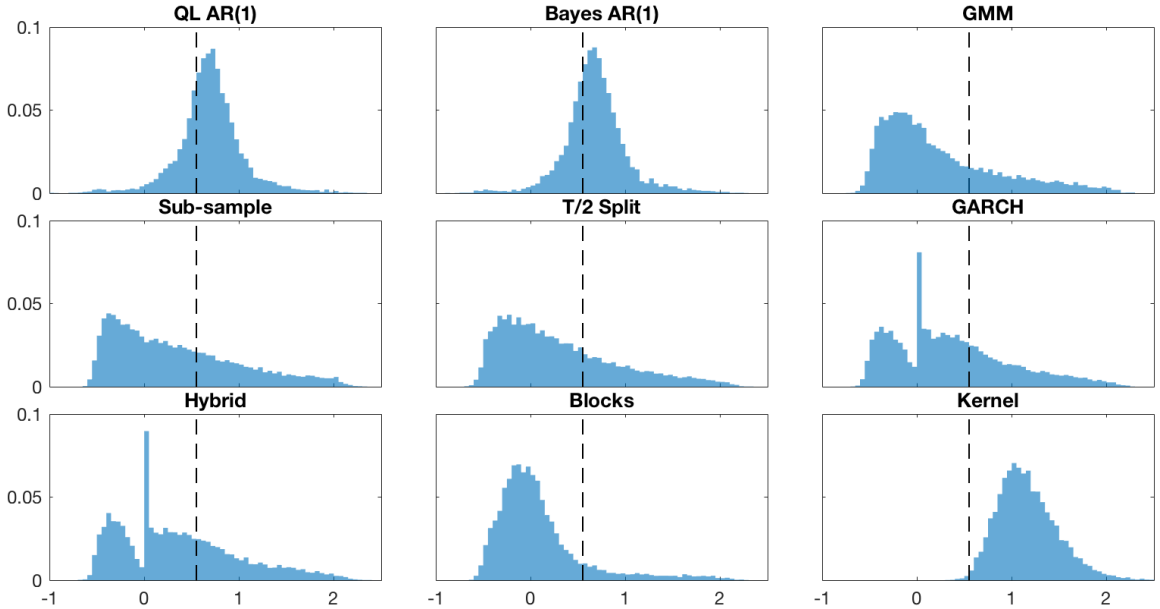
(b) Distribution of H_{12} estimates



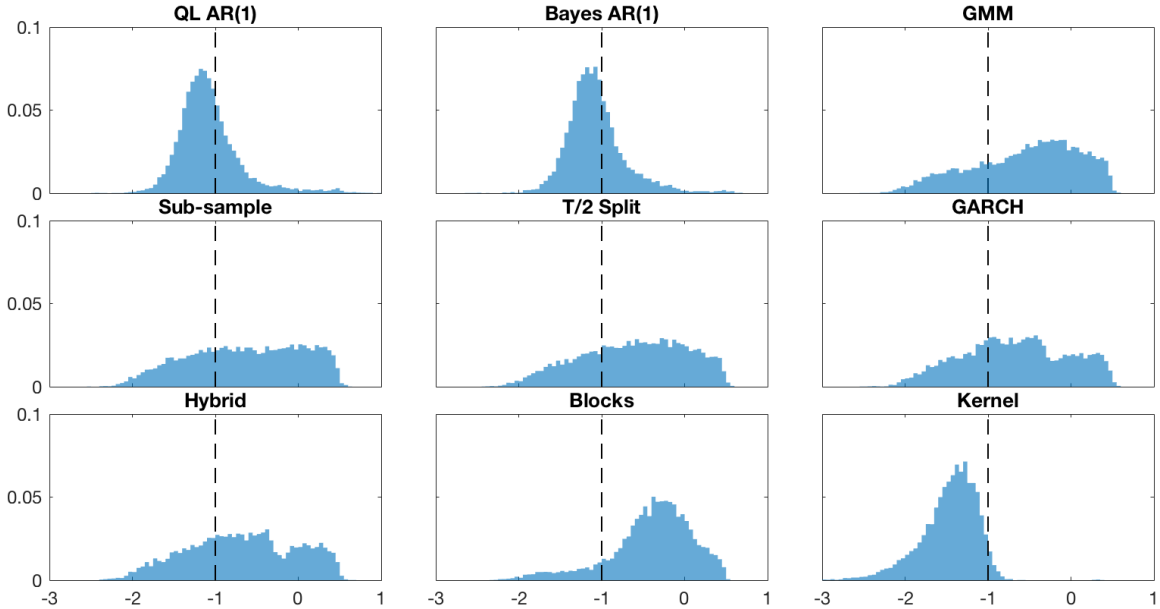
Distribution of estimates of H_{21} and H_{12} for various estimators for "weak" GARCH(1,1) DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 30: Distribution of estimates: empirically calibrated Markov Switching, $T = 200$

(a) Distribution of H_{21} estimates



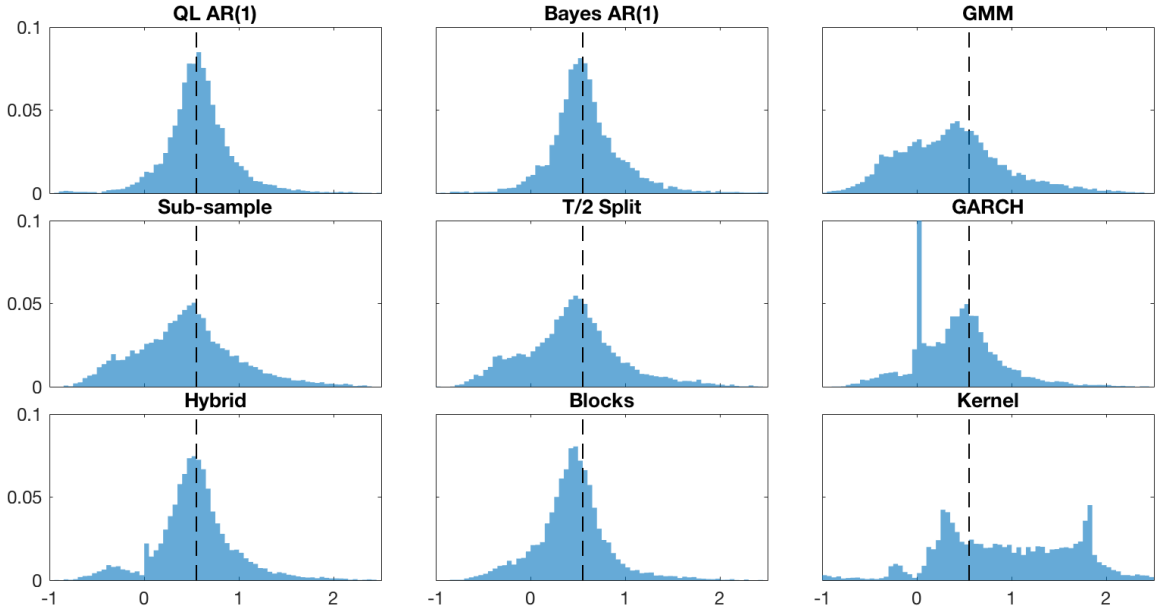
(b) Distribution of H_{12} estimates



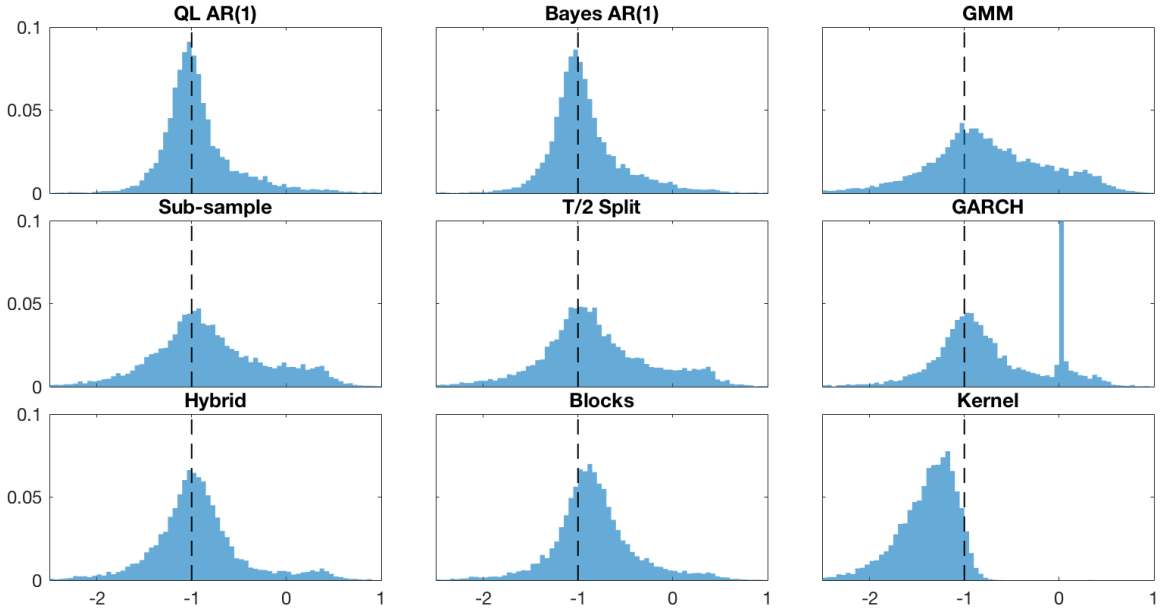
Distribution of estimates of H_{21} and H_{12} for various estimators for Markov switching DGP, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 31: Distribution of estimates: empirically calibrated AR(1) SV, $T = 100$

(a) Distribution of H_{21} estimates



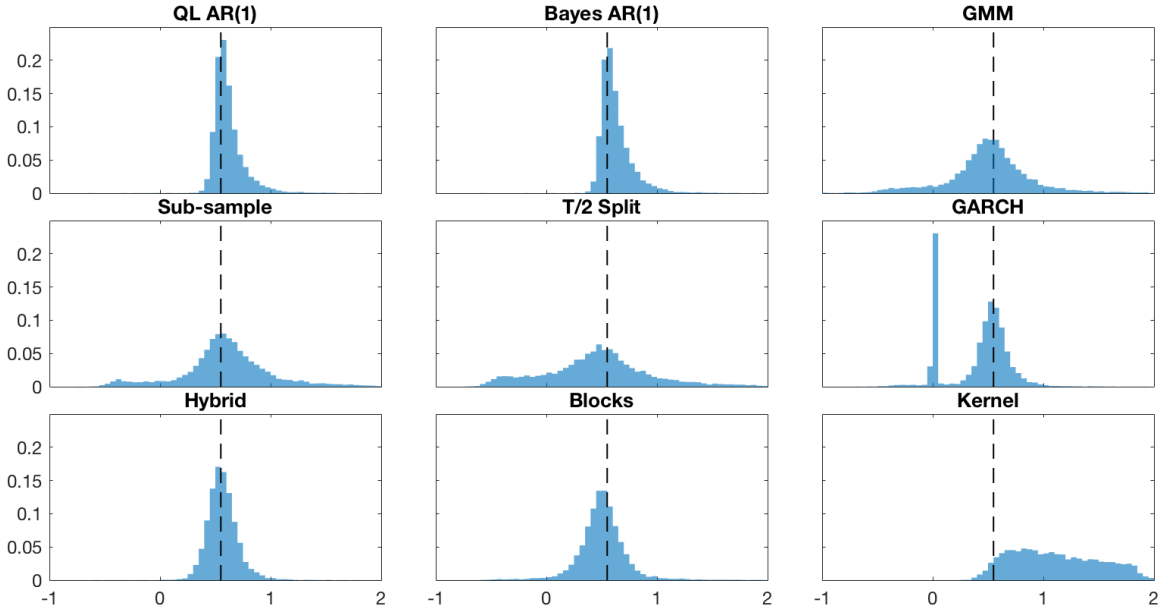
(b) Distribution of H_{12} estimates



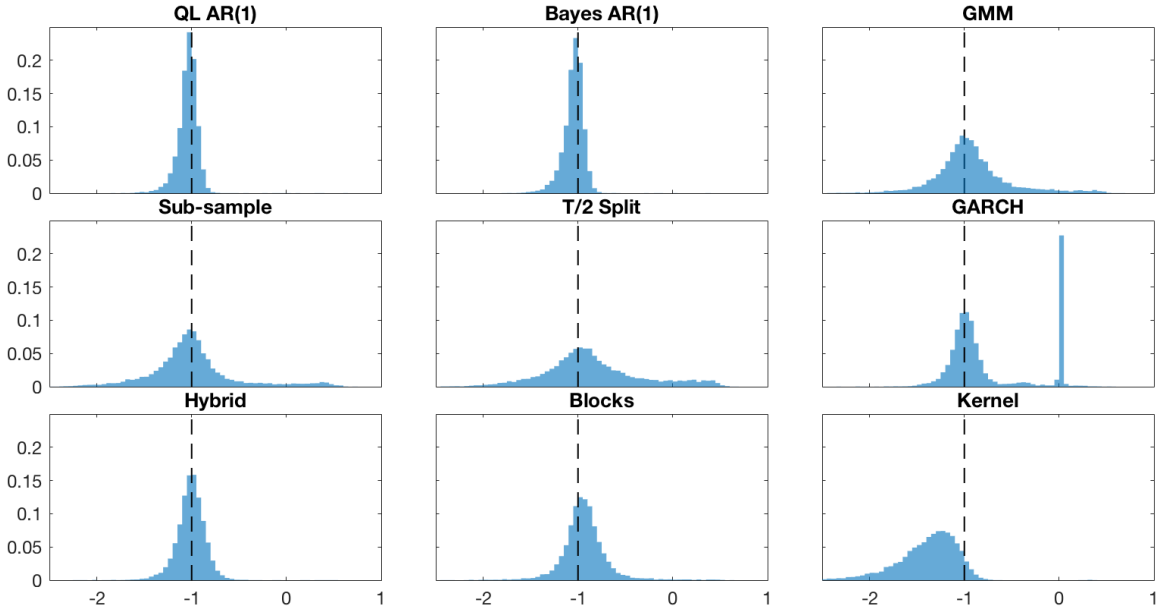
Distribution of estimates of H_{21} and H_{12} for various estimators for empirically-calibrated AR(1) SV DGP, $T = 100$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm. Note that the peak of the GARCH distribution is truncated to allow all estimators to be viewed on standardized axes.

Figure 32: Distribution of estimates: empirically calibrated AR(1) SV, $T = 500$

(a) Distribution of H_{21} estimates



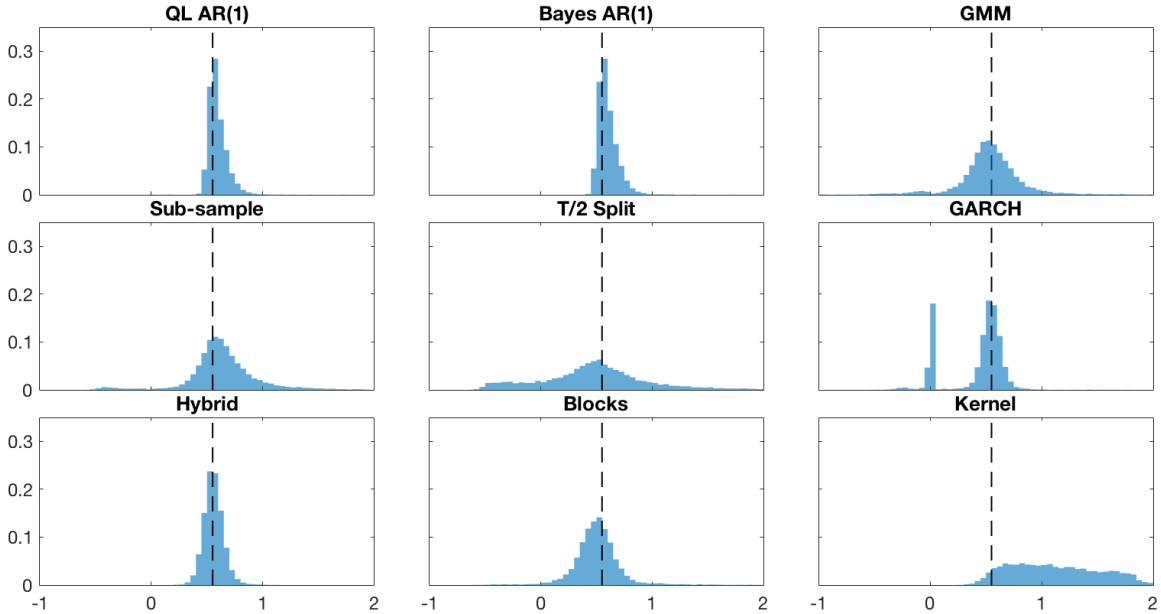
(b) Distribution of H_{12} estimates



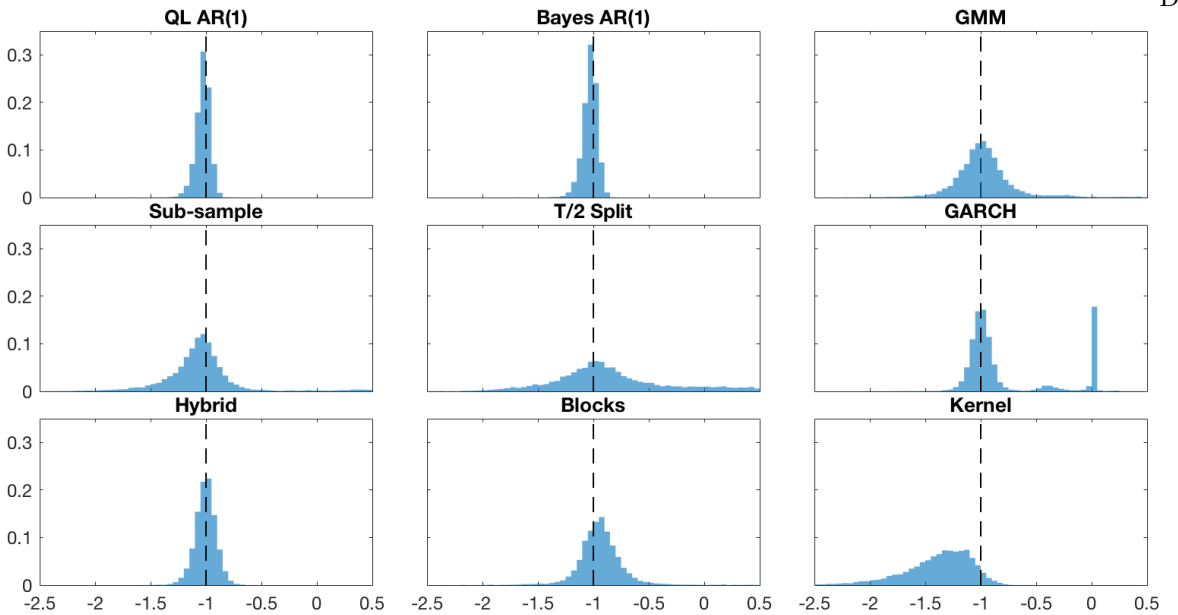
Distribution of estimates of H_{21} and H_{12} for various estimators for empirically-calibrated AR(1) SV DGP, $T = 500$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 33: Distribution of estimates: empirically calibrated AR(1) SV, $T = 1000$

(a) Distribution of H_{21} estimates



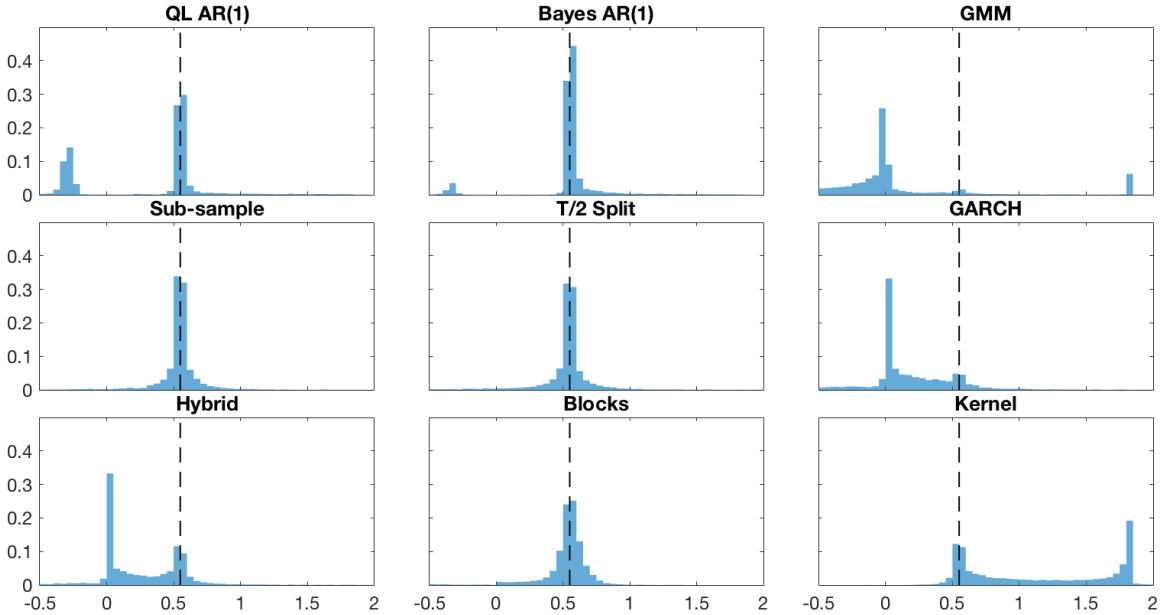
(b) Distribution of H_{12} estimates



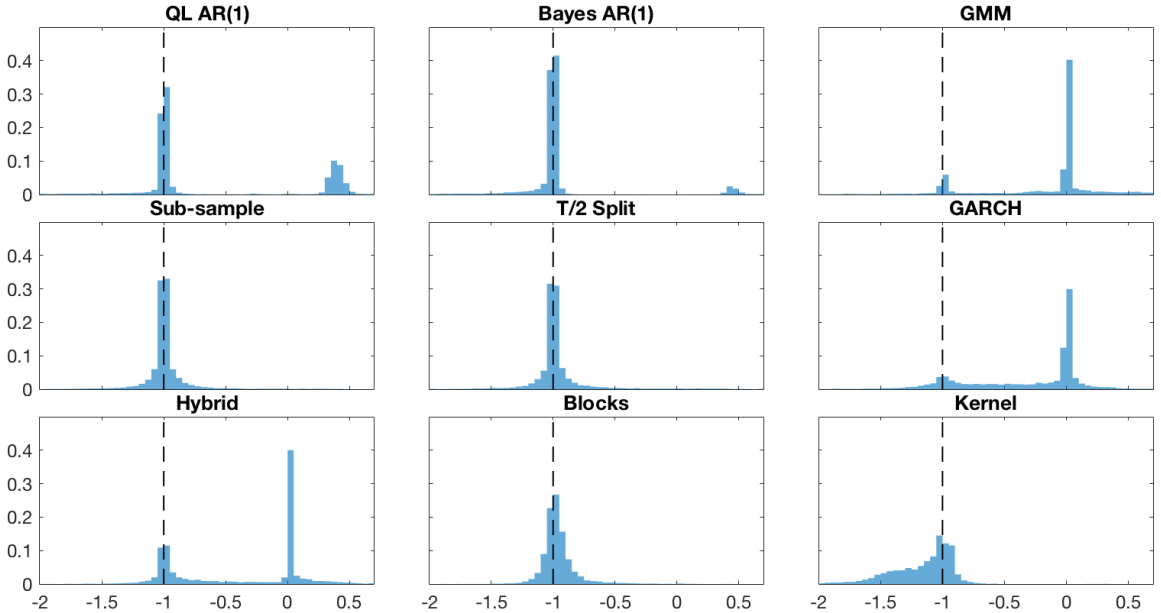
Distribution of estimates of H_{21} andH_{12} for various estimators for empirically-calibrated AR(1) SV DGP, $T = 1000$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

Figure 34: Distribution of estimates: empirically calibrated AR(1) SV, $T = 200$, t_1 shocks

(a) Distribution of H_{21} estimates



(b) Distribution of H_{12} estimates



Distribution of estimates of H_{21} and H_{12} for various estimators for empirically-calibrated AR(1) SV DGP, t_1 shocks, $T = 200$, 10,000 draws. Details of estimators can be found in Table 4. Labeling proceeds via an infeasible method matching H estimates to the true H to minimize L_2 norm.

D.3 Empirical results for alternative estimators

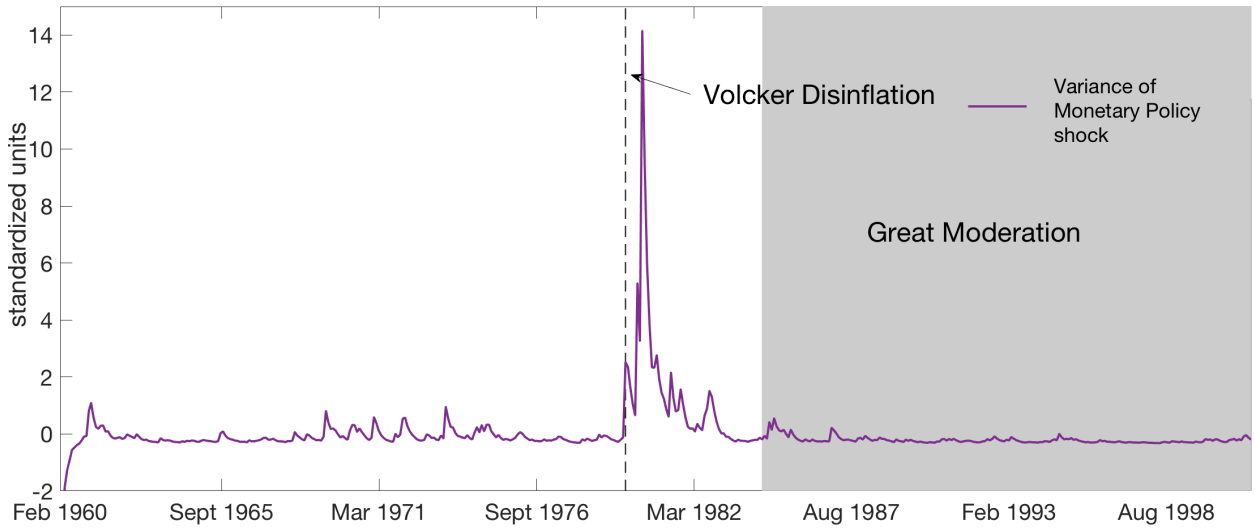
As a robustness check, the empirical application is further estimated using the hybrid GARCH estimator. The results are very close to those of the quasi-likelihood AR(1) estimator reported in the text. Estimation proceeded via maximum likelihood, running Matlab's global search from many different start values to obtain a "global" minimum. Table 14 reports the parameter estimates. The estimates themselves are quite different from those resulting from the AR(1) log SV form; however, there are local minima in the neighbourhood of the previous results. The Cholesky structure is strongly rejected by the joint test. Figure 35 shows the variance path for the monetary policy shock series. Again, it accords with conventional wisdom of the scale of monetary policy shocks over the past decades. Figure 36 compares the three key IRFs to those obtained by BBE and the quasi-likelihood AR(1) SV estimates. Despite the difference in point estimates of H, the response paths remain similar; the overall conclusions are robust to the functional form used. Finally, Figure 37 reports the IRFs for the full collection of macro variables. Here, the standard errors are constructed by estimating the covariance matrix of the reduced-form IRF via the bootstrap method described in Appendix B.4, computing the covariance of H estimates using the QML methods described in the text, forming a block diagonal matrix, and using a first-order approximation.

Table 14: Hybrid GARCH estimates of $H^{(FFR)}$

	Hybrid	QL AR(1) SV
Factor 1	-0.14	0.27
	(0.19)	(0.25)
Factor 2	0.22***	0.50***
	(0.05)	(0.14)
Factor 3	-0.69***	-0.28**
	(0.10)	(0.13)
FFR	1	1
Joint test	1268.10***($p < 0.001$)	75.26*** ($p < 0.001$)

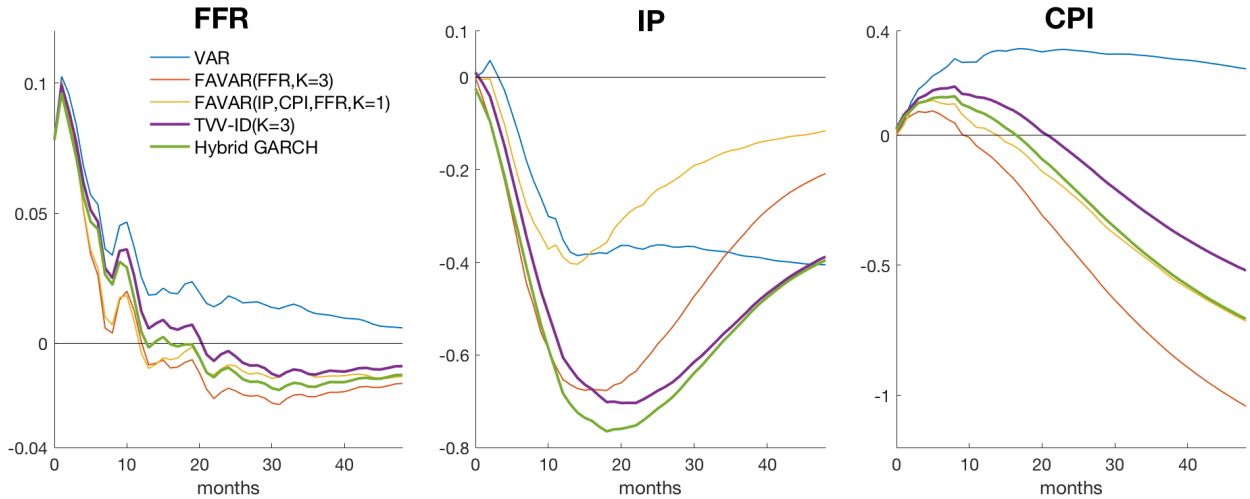
Comparison of the monetary policy column of H under TVV-ID estimated via the hybrid GARCH and the quasi-likelihood AR(1) SV. Labeling proceeds using the zeroth order FEVD method. Care must be taken in interpreting the scale of the factor coefficients as these are arbitrarily scaled factors, although they are standardized and estimated based on standardized macro time series. The joint Wald tests are based on the QML asymptotic variance and the Müller (2013) covariance estimator.

Figure 35: Hybrid variance path



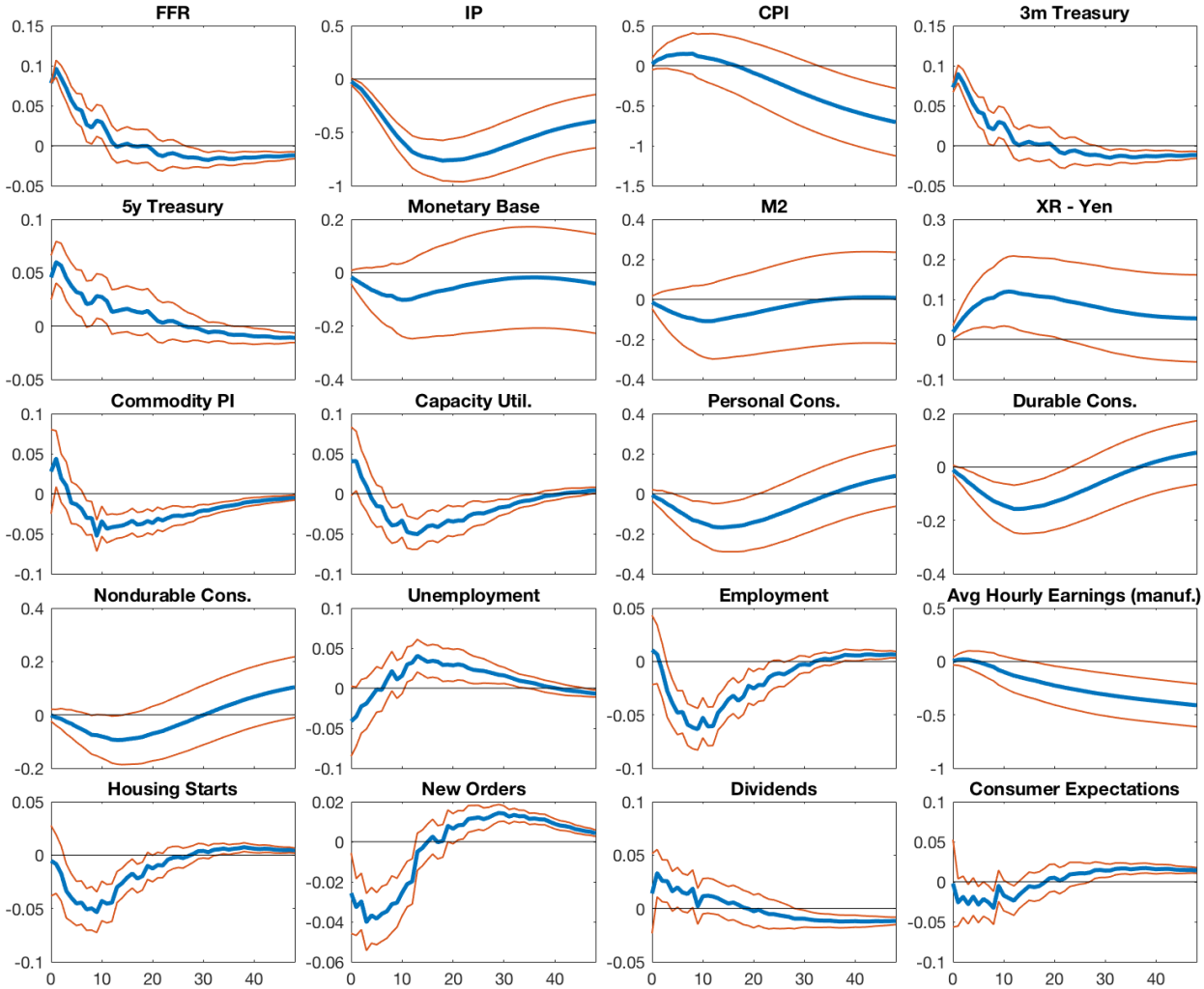
Variance of the monetary policy shock identified via TVV-ID estimated using the hybrid GARCH. Labeling proceeds using the zeroth order FEVD method. Volcker period and Great Moderation added for reference.

Figure 36: Cholesky vs TVV-ID (with Hybrid)



Impulse responses to a 25 basis point shock to the Fed Funds rate, at a monthly frequency, based on the results of Table 14 and BBE. The vertical axis is in standard deviations. TVV-ID for quasi-likelihood AR(1) results in purple and TVV-ID via hybrid GARCH in green; additional lines replicated from BBE Figure I for comparison.

Figure 37: Impulse responses for important macro variables (Hybrid)

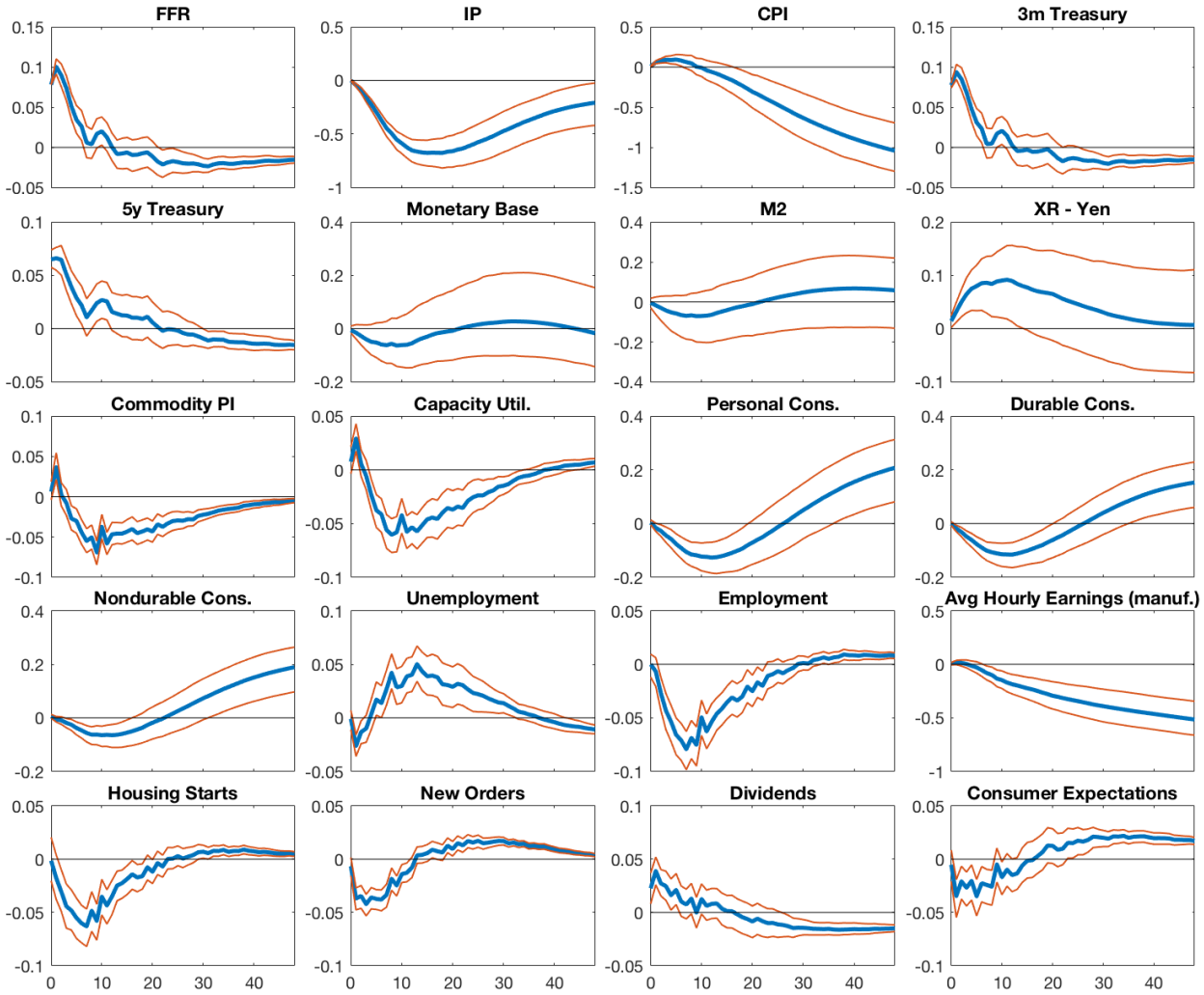


Responses of 20 key macroeconomic variables to a 25 basis point Fed Funds shock identified via TVV-ID estimated by hybrid GARCH, based on the results reported in Table 14. The vertical axis is in standard deviations. Compare to Figure 10. 95% confidence intervals are computed using the method described in Appendix B.4.

D.4 Additional empirical results

Figure 38 replicates the Figure II IRFs of BBE with the addition of confidence intervals computed via the Gospodinov & Ng (2013) method adopted in this paper, for ease of comparison with the TVV-ID results.

Figure 38: Impulse responses for important macro variables: Cholesky



Impulse responses to a 25 basis point Fed Funds shock for key macro variables based on Cholesky decomposition. The vertical axis is in standard deviations. This is a replication of BBE Figure II, with 95% confidence intervals using the method described in Appendix B.4, to allow better comparison with Figure 10.