Identifying Shocks via Time-Varying Volatility*

Daniel J. Lewis

May 2 2018

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 new identification method that identifies the SVAR up to $n!$ shock orderings using only moment equations implied by an arbitrary stochastic process for the variance. Unlike previous work, this result requires only weak technical conditions. In particular, it requires neither parametric assumptions nor the specification of variance regimes. I propose intuitive criteria to select among the orderings and show that this selection does not impact inference asymptotically. As an empirical illustration, I consider Kilian's (2009) work examining oil prices and their macroeconomic effects. This exercise strengthens his results by failing to reject his lower-triangular assumption and replicating his macroeconomic conclusions.

Keywords: identification, impulse response function, latent variables, structural shocks, time-varying volatility, Cholesky decomposition, oil price shocks.

JEL Codes: C32, C38, C58, E31, E32, Q41, Q43.

*Helpful comments provided by Jim Stock, Gabriel Chodorow-Reich, Neil Shephard, Mikkel Plagborg-Møller, Lutz Kilian, Jón Steinsson, Giri Parameswaran, Domenico Giannone, Andrew Patton, and various seminar participants.

1 Introduction

A central challenge in structural vector autoregression (SVAR) analysis is identifying 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 endogenous 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 Kilian & Lütkepohl (2017) 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 smaller literature that considers identification based on 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. More recent work has considered Markov switching (Lanne, Lütkepohl, & Maciejowska (2010)) and allowed for smooth transitions between regimes (Lütkepohl & Netšunajev (2015)). Little consideration has been given to what happens when these parametric structures 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 (unspecified) form, identification follows from the autocovariance of the volatility process. As a result, unlike previous approaches, it accommodates both conditional and unconditional heteroskedasticity. 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), Hyvärinen et al (2010), Lanne, Meitz, & Saikonnen (2017), and Lanne & Lütkepohl (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 autocovariance of the volatility process satisfy a weak rank condition.

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. In many cases, such conventional assumptions can also be tested as overidentifying restrictions, which may increase the economic interpretability of the statistically-recovered shocks (Kilian & Lütkepohl (2017)). Significantly, it is transparent to discuss the impact that such assumptions have on the ultimate estimates of the relevant elements of the response matrix.

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.

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 a simulation study. Likelihood-based approaches perform well in this context, even under misspecification, akin to known results of Quasi-Maximum Likelihood. Thus, TVV-ID is a more reliable option for researchers not possessing substantive information about the underlying volatility process.

As an empirical illustration, I apply TVV-ID to Kilian's (2009) work on oil shocks. Lütkepohl & Netšunajev (2014) consider this paper as a test case in their exploration of identification using Markov switching models as a means to test conventional identification restrictions, motivating the choice to examine it here. Kilian (2009) recovers three shocks impacting oil prices from an SVAR with recursive structural assumptions: an oil supply shock, an aggregate demand shock, and an oil-specific demand shock. He finds that the source of oil price movements has important implications for their effects on the U.S. macroeconomy. While these identifying assumptions are plausible, I replicate his analysis, but use TVV-ID instead of recursive assumptions to identify the structural shocks. This means I can test his recursive identifying assumptions as overidentifying restrictions; I am unable to reject his lower triangular structure, and in the leading implementation, I estimate precise zeros where he assumes them. This generalizes the results of Lütkepohl & Netšunajev (2014), who are unable to reject both the recursive structure and the assumption that H is fixed, based on a Markov switching model. The macroeconomic impact of these shocks is likewise virtually unchanged.

I overcome the main scope for challenge to Kilian (2009) by validating his identifying assumptions and thereby strengthen his results. In so doing, I show that his conclusions are robust to different sources of identification.

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 & Lütkepohl (2008), Ehrmann, Fratzscher, & Rigobon (2011), Eichengreen & Panizza (2016), Ehrmann & Fratzscher (2017), and Hébert & Schreger (2017). Normandin & Phaneuf (2004), Normandin (2004), Doz & Renault (2004), and Lütkepohl & Milunovich (2016), amongst others, have followed path-based identification, mostly in monetary economics and finance. While many applications come from such fields, 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. On the other hand, macro models with time-varying volatility have been estimated for some time, without exploiting its implications for identification. A lower triangular structure on H has generally been retained due to the belief that doing so is necessary for identification (e.g. Primiceri (2005)). This literature offers an immediate avenue for the exploitation of my results.

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 theory and simulation studies. The empirical illustration to Kilian (2009) follows in Section 6. Section 7 concludes. Proofs can be found in Appendix A; additional details on the application of TVV-ID, of interest to practitioners, are found in Appendix B. Discussion of further estimation approaches and full simulation results are available in the Online Appendix.

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 i^{th} row of matrix A; $A^{(j)}$ denotes the j^{th} column of matrix A; A_{ij} denotes the ij^{th} element of matrix A; $A^{(-i)}$ denotes all columns of A except for the i^{th}, 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 $\{x_1, x_2, \ldots, x_t\}$; $E_t\,[\cdot]$ denotes a

time-specific expectation, i.e. the mean value of x_t at time t, as opposed to across t, and similarly for $E_{t,s}[\cdot]$ when both time t, s variables are contained in the argument.[1]

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. More broadly, this represents a general decomposition problem. 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 how 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. *Shocks satisfy conditional mean independence,* $E[\varepsilon_{it} \mid \varepsilon_{-is}] = 0$ *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 slightly 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

[1] This notation is used to make explicit that stationarity is not being assumed, unless otherwise noted, and to avoid the ambiguity (and possible non-existence) present in simply writing $E[x_t]$ in a non-stationary context. The use of E_t should not be confused with reference to the t information set; when a specific information set is intended, I condition on it explicitly.

the overall covariance of η_t, $E\left[\eta_t \eta_t'\right] = \Sigma_\eta$. However, it is well-known that these equations can only identify H up to an orthogonal rotation, Φ ($\Phi\Phi' = I$). Observe

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

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 $n\left(n+1\right)/2$ equations, but there are n^2 unknowns. This is the fundamental identification problem posed by the SVAR methodology and indeed many related models (e.g. factor models).

Variation in Σ_t may allow the researcher to overcome the limitations of (2). 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 (2018)). Without loss of generality, assume σ_{2t}^2 is the variance that changes, while $\sigma_{1t}^2 \equiv \sigma_1^2$, constant. Denote

$$\sigma_t^2 = \begin{bmatrix} \sigma_1^2 \\ \sigma_{2t}^2 \end{bmatrix}, \; H = \begin{bmatrix} 1 & H_{12} \\ H_{21} & 1 \end{bmatrix}.$$

The conditional variances of the reduced-form innovations are given by $E_t\left[\eta_t\eta_t' \mid \sigma_t\right] = H\Sigma_t H'$. Given two subsamples, A, B, containing the sets of time points T_A, T_B, Appendix A (and prior work) shows that

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

where the $\Delta\left(\cdot\right)$ operator represents the difference in expectation of the argument between subsamples T_A, T_B. Assuming that $\Delta\left(\sigma_{2t}^2\right) \neq 0$, H_{12} can thus be identified in closed form. Fourth order stationarity is the only assumption made on the form of the stochastic process for σ_{2t}. 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 is indeed fixed. If there are in fact regimes, they need not be known or correctly specified, as noted in Rigobon (2003). However, if the value of the σ_{2t} process is in fact constant, $\Delta\left(\sigma_{2t}^2\right)$ would be zero in population, and identification fails.[2]

Stated this way, Rigobon's approach provides moment conditions based on means of the variance process, which can yield identification for many processes, but, there is no reason not to consider alternative arguments based on other moments. In each period, there is motivation for an instrumental

[2]In this case, if regimes are instead estimated from the values of η_t, the resulting estimates of $\Delta\left(\sigma_{2t}^2\right)$ are not zero in population since differing regimes are driven by realized shock values, but this source of variation results in bias, as discussed in Section 5.

variables approach. Noting

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

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

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

by Assumption 0.4 and the fact that σ_1 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 $\eta_{2t}\eta_{1t}$, the 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 assumptions on the stochastic process σ_{2t} is that it is fourth-order stationary (for expositional simplicity) and that $E\left[\varepsilon_{2t}^4\right] < \infty$. 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 p^{th} 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 around the regime break dates. In a stochastic volatility (SV) process, the condition holds provided some AR coefficient of the variance is non-zero. In a Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) model, provided at least one of the auto-regressive parameters is non-zero, it will likewise hold.[3] The condition can be verified for other stochastic processes 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. I have made no assumptions about whether the heteroskedasticity is conditional or unconditional (either implies a suitable autocovariance), unlike in the Rigobon approach that presumes unconditional heteroskedasticity. As noted above, and discussed in detail in Section 5,

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

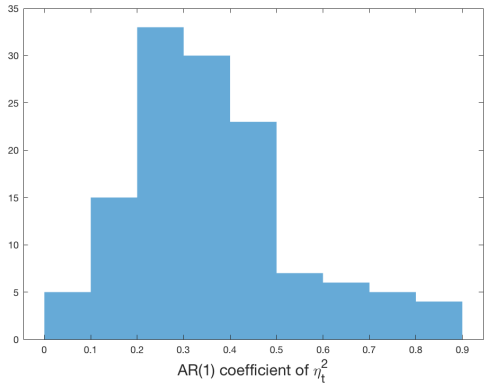

Figure 1: *Distribution of estimated AR(1) coefficients of η_t^2. Time series η_t are obtained as reduced-form innovations from AR(12) processes fitted to each of McCracken & Ng's 128 FRED-MD monthly time series. The figure displays the distribution of the implied AR(1) coefficients of η_t^2.*

fitting regimes to apply the Rigobon approach runs the risk of substantial bias if the variances in fact evolve continuously, and thus much of the regime determination is driven by the draws of the disturbances. In Sentana & Fiorentini's (2001) work, identification is robust only to a very small set of specification errors relative to the imposed GARCH functional form; this is illustrated in the simulation study, Section 5. In contrast, in this example, I have required that the stochastic process is stationary and exhibits some degree of persistence.

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(12) process for many macro time series. I consider the 128 monthly series spanning 1959-2018 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 hypothesis of zero autocovariance for 99 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 118 of the series, the 5% level for 113, and the 1% level for 103. Thus, it appears that the identifying condition is 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

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

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 are made on dimension and the stochastic process of σ_t to yield an elegant closed-form solution; I now relax those restrictions to yield a much more general result.

2.2 Identification via time-varying volatility

Identification via time-varying volatility applies under much weaker conditions than those outlined to build intuition above. Again, let

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

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

Assumption 1. *For every* $t = 1, 2, \dots, T$,

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

In addition, I make a preliminary assumption on H:

Assumption 2. H is time-invariant.

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 $\varepsilon_1, \dots \varepsilon_{t-1}$), amongst others (or, more broadly, both unconditional and conditional heteroskedasticity).[4]

Decomposition

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

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

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

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).[5] Then

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

The simplification from (5) to (6) in the first term is surprising and follows due to the diagonality of Σ_t using Assumption 1.2. This feature plays a key role in properties established later. From the definition of V_t, Assumptions 1.1, 1.3, and 2, $E_t\left[V_t \mid \sigma_t, \mathcal{F}_{t-1}\right] = 0$, 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 Assumption 1.3 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 3 expands on 1.3 to allow the establishment of useful properties of $\zeta_t = vech\left(\eta_t\eta_t'\right)$.

Assumption 3. *For every t,*

1. $\mathrm{Var}_t\left(\sigma_t^2\right) < \infty$,
2. $\mathrm{Var}_t\left(\varepsilon_t\varepsilon_t'\right) < \infty$.

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

Proposition 1. *Under Assumptions 1.1-1.2, 2, & 3,*

$$Cov_{t,s}\left(\zeta_t, \zeta_s\right) = L\left(H\otimes H\right) GM_{t,s}\left(H\otimes H\right)' L',\ t > s \quad (7)$$

where

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

This equation has the desired form: it represents an observable quantity, $Cov_{t,s}\left(\zeta_t, \zeta_s\right)$, as a product of H and the $n\times n^2$ $M_{t,s}$ 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:

[5] 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 $d = diag\left(D\right)$.

Assumption 4. *For $t > s$, $E_{t,s}\left[\sigma_{it}^2\left(\varepsilon_s\varepsilon_s^{'} - \Sigma_s\right)\right]$ is diagonal for all $i = 1, 2, \ldots, n$.*

This means that ARCH effects to σ_t^2 cannot depend on any off-diagonal elements of $\varepsilon_s\varepsilon_s^{'}$. Assumption 4 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.[6] I can adapt Proposition 1 under Assumption 4:

Proposition 2. *Under Assumptions 1.1-1.2, 2, 3, & 4, (7) simplifies to*

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

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

$\check{M}_{t,s}$ subsequently simplifies further to $Cov_{t,s}\left(\sigma_t^2, \sigma_s^2\right)$ if the process exhibits no conditional heteroskedasticity. Significantly, the imposition of Assumption 4 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 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 strengthen the assumptions on H from Assumption 2:

Assumption 2′. *H is time-invariant, invertible, with a unit diagonal.*[7]

Given Assumption 2′, the conditions under which equation (7) yields a unique solution for H are established by Theorem 1.

Theorem 1. *Under Assumptions 1.1-2, 2′, & 3, equation (7) holds. Then H and $M_{t,s}$ are jointly uniquely determined from (7) (up to labeling of shocks) provided $rank(M_{t,s}) \geq 2$ and $M_{t,s}$ has no scalar multiple rows.*

[6] 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 4.

[7] The unit diagonal assumption is a normalization, without loss of generality.

In the case where Assumption 4 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).

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. 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 (Lanne & Lütkepohl (2010), amongst others), Sims' implementation of subsample identification (see Section 5), or identification in finite mixture models, and is discussed in Chapter 14 of Kilian & Lütkepohl (2017). In some cases, the labeling of shocks is unnecessary (as in factor models) – and identification is complete. This is not the case for policy analysis; labeling is discussed in detail 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.

The conditions of Theorem 1 impose interpretable mild restrictions on the process σ_t^2. The rank condition 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. Recall that, in the SVAR setting, 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. These points are illustrated in the following example.

Example. 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, then, assuming stationarity, the autocovariance matrix of σ_t^2 has the form

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

where a is some scalar. In this case, identification is sought from

$$L\left(H \otimes H\right) Ga \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} G' \left(H \otimes H\right)' L'.$$

This matrix can be re-written as

$$vech\left(a^{1/2}HI_2H'\right)vech\left(a^{1/2}HI_2H'\right)' = a \times vech\left(HH'\right)vech\left(HH'\right)'$$

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 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 nonlinear function $r\left(\cdot\right)$,

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

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.

When $n > 2$, beyond the rank condition, a scalar multiple condition applies to the rows of 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 guaranteeing uniqueness break down. In practice, there is little reason to think this condition will be violated; rather, it is more likely to lead to a weak identification problem if nearly violated. For a discussion of weak identification in TVV-ID, see Appendix B.2. Nevertheless, in some finance settings, see 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 1.1-1.2, 2′, & 3, equation (7) holds. Then $H^{(j)}$ is identified from* (7) *provided $rank\left(M_{t,s}\right) \geq 2$ and $M_{t,s}$ contains no rows proportional to row j.*

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 4 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

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

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

Theorem 2. *Under Assumptions 1.1-2, 2′, & 3, equation (7) holds. Then H is uniquely determined from (7) and (10) (up to labeling of shocks) provided $\left[\begin{array}{cc} M_{t,s} & E_t\left[\sigma_t^2\right] \end{array}\right]$ has rank ≥ 2 and no scalar*

multiple rows.

Again, a symmetric extension applies under Assumption 4. Theorem 2 requires that, in order for identification to fail, a scalar multiple assumption must also relate $E_t\left[\sigma_t^2\right]$ 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.*

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 2′

Even for $n = 2$, the system of equations is overidentified, with the degree of overidentification increasing in n. This means tests exploiting overidentification can be conducted. This is an advantage over many identification approaches in this setting, where strong assumptions are required to yield even 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, (e.g. Stock & Watson (2017), Plagborg-Møller (2018)). 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 (the recent work of Chahrour & Jurado (2017) discusses some exceptions). Thus, a test indicating misspecification, like a $J-$test, 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, H remains fixed. While this may seem inconsistent, there are several points to consider. First, no existing identification scheme flexibly handles time-varying H (Carreiro, Clark & Marcellino (2018) do so under a very specific functional form). Even simple identification based on Cholesky structure, 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. Since the Rigobon approach already works 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 economic 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. 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.[8]

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 $Cov_{t,s}\left(\zeta_t, \zeta_s\right)$ is known for any t, s pair, the identification argument holds. However, stationarity can play a role in estimation, where assumptions are required for $Cov_{t,s}\left(\zeta_t, \zeta_s\right)$ 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 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 by 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. $E\left[vec\left(\eta_t\eta_t'\right) vec\left(\eta_{t-p}\eta_{t-p}'\right)'\right]$.

[8]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.

2.3 Nesting the existing literature

TVV-ID holds in virtually any case where previously developed identification schemes apply. While Proposition 4 of Sentana & Fiorentini (2001) shows 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. Sentana & Fiorentini's ability to apply their 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, applying their result requires a one-to-one mapping between $\left(\eta_t, H, \sigma_1^2\right)$ and $\sigma_{2:T}^2$ (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.[9] 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 itself is literally dependent on such an assumption, or some other restrictive method to map $\left(\eta_t, H, \sigma_1^2\right)$ to $\sigma_{2:T}^2$ one-to-one. TVV-ID clearly nests the GARCH-based identification of Sentana & Fiorentini (2001) – a GARCH(1,1) 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_t^2 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 may fail asymptotically.

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

When there are 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, despite extensions like Lütkepohl & Netšunajev (2017) to allow for smooth transitions between fixed regimes. External information can convincingly isolate periods of high and low volatility, as in the original Rigobon (2003) paper, Rigobon & Sack (2004), and Lanne & Lütkepohl (2008), up to more recent papers such as Nakamura & Steinsson (2018) and Brunnermeier et al (2017). These studies make arguments such as "the volatility of the monetary policy shock will be higher on monetary policy 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 or more 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 or Markov-switching model, (see e.g. Lanne et al (2010)). 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 inherently be endogenous, including the possibility of substantial bias, are discussed in Section 5. TVV-ID does not face these difficulties, as no variance path must be estimated.

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. Kilian and Lütkepohl (2017) discuss how there may in fact be some difficulty in interpreting these as economically meaningful shocks, given the purely statistical methods used to derive them; this step helps to develop such interpretations. In this section, I first outline 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 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 an identified point. The same problem arises in the identification schemes based on non-Gaussianity, and is discussed in some detail in Kilian and Lütkepohl (2017). In the non-Gaussianity literature, Lanne & Lütkepohl (2010) make restrictions on the H matrix to label monetary policy shocks 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. Hyvärinen et al (2010) and Gouriéroux & Monfort (2014) do not address the issue. Ludvigson, Ma & 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. Some sets of assumptions eliminate candidates from the identified set, while others select a best candidate. It is notable that for virtually any traditional macroeconomic identification scheme one would otherwise be forced to use to identify a SVAR, there is a weaker analog that can be used here to choose between the shock series furnished by TVV-ID. 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.

It is important to note that some work has considered the problem of interpreting the shocks recovered using statistical identification methods (like identification via heteroskedasticity) as a more difficult problem. Kilian & Lütkepohl (2017) argue that these shocks need not be economically meaningful. The labeling exercise outlined above does not, however, necessarily assume the shocks are meaningful - it is possible that no shock satisfies a theoretically-motivated labeling criterion. Nevertheless, a researcher holding the concerns of Kilian & Lütkepohl (2017) can consider, as those authors suggest, whether a statistically-recovered shock represents an economic shock by formally testing conventional identifying assumptions as overidentifying restrictions. If the restrictions cannot be rejected, the shock satisfies the theoretical properties of a conventional SVAR shock. Significantly, while conventional restrictions are being tested at this post-estimation stage, the recovery of shocks and responses is unrestricted and more flexible than in standard approaches. Such exercises do, however, still generally require an initial labeling to determine a shock series or column of H to compare to the theoretical restrictions. A second alternative is a more informal approach, evaluating the extent to which the impulse response functions (IRFs) are compared to those based on economic theory (or conventional models identified using such theory), as in Brunnermeier et al (2018) or Lütkepohl & Netšunajev (2014).[10] There is clearly scope to develop more rigorous frameworks to conceptualize the shocks recovered from statistical identification approaches like TVVID, conventional methods using heteroskedasticity, and non-Gaussianity. The underlying point is that TVVID makes progress on existing approaches by delivering the candidate shocks under weak assumptions.

Regardless of whether the shocks recovered are assumed to be meaningful, or more skepticism is taken in presuming interpretability, labeling then plays a role either in isolating the shock of interest directly, or furnishing appropriate shocks to be tested against structural assumptions. A collection of possible approaches is outlined below.

- In a context where a lower-triangular assumption could otherwise be used and is considered plausible, the columns can be ordered so as to come closest to the zero restrictions under some norm. This lets the data dictate more realistic near-zeros instead of assuming sharp zeros. A similar analog exists for Uhlig's (2005) sign restrictions, or Blanchard & Quah's (1989)

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

long-run restrictions.

- Imposing a restriction on a column of interest will ex ante label that column. Conversely, restricting all other columns has the same effect. Such assumptions can be arbitrarily sparse - there is no need for a fully lower-triangular structure. This approach is adopted in Lanne & Lütkepohl (2010).

- As is common in the Rigobon approach, assumptions on how the shock variances change (perhaps in conjunction with historical episodes) can label the columns.

- A forecast error variance decomposition-type approach can label H by supposing that within a period, the majority of unpredictable variation in a particular series is driven by a certain type of shock.

- If there is an external instrument available for a policy shock, as described by Stock (2008), it can be used to select the shock series that is best correlated with it, instead of making a strict exogeneity assumption.

- Certain magnitudes of responses can be ruled out as implausible. Even very loose magnitude restrictions can be helpful in contexts where the estimated columns have drastic differences in relative magnitudes (a wrong normalization inflates elements dramatically).

- Plotting IRFs for the recovered shocks and attempting to name the shocks based on the dynamics is also an option, as in Brunnermeier et al (2017).

A more detailed discussion, complete with examples, is presented in Appendix B.1.

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" mentioned 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. Section C.1 of the Online Appendix discusses implementation of the Least Mean Squares filter in this context, which can furnish volatility paths under minimal parametric assumptions.

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 $\widehat{H^{(j)}}$, 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). In the context considered here, the strong notion of consistency of the labeling criterion makes it direct to show that 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 $\widehat{H^{(j)}}$ will be that of the j^{th} column 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 more subjective "economic" identifying assumptions are only used to label a defined set of shocks, or, equivalently, to identify which column $H^{(j)}$ 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 labeling and thus 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 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.[11]

[11] 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 1, 2′, & 3) have not been enforced prior to this labeling problem, so it is possible to leave all "economic" decisions in the hands of the reader.

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. Estimation using a GMM approach follows directly from the identifying equations. In particular, with the addition of a stationarity assumption, Equation (7) can be consistently estimated, and standard GMM results apply. However, in practice, (and in the simulation study) there may be challenges to this approach, which are discussed in Appendix B.3. In this section, I consider other approaches: I outline likelihood-based inference via Markov Chain Monte Carlo (MCMC) and describe what I call hybrid GARCH. Further options exist, including 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.3. 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 variances; to some extent this may seem a return to parametric assumptions this paper set out to avoid. However, thanks to the general identification arguments offered above, *identification* is not tied to a particular functional form. A researcher can specify any functional form for time-varying volatility provided it implies an autocovariance. In analogy to a simple IV model, determining the validity of an instrument (identification) is a separate problem from deciding whether to estimate via two-stage least squares or choosing a parametric model for maximum likelihood. It is also possible to fit multiple functional 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[12]. For a dimension i, it takes the form

$$\log\left(\sigma_{it}^2\right) = \mu_i\left(1-\phi_i\right) + \phi_i \log\left(\sigma_{it-1}^2\right) + e_{it}, \tag{11}$$

where e_{it} and e_{jt} can have arbitrary covariance for $i, j \in 1, ..., n$. 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.3.

[12]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 Engle (2001), for example. Consider the GARCH(1,1) functional form for $t = 2, 3, \ldots, T$:

$$\sigma_{it}^2 = \mu_i \left(1 - \bar{\psi} - \bar{\Upsilon}\right) + \bar{\psi}\sigma_{i,t-1}^2 + \bar{\Upsilon}\varepsilon_{i,t-1}^2, i = 1, 2, \ldots, n, \text{ and } \mu_i, \bar{\psi}, \bar{\Upsilon} \geq 0, \tag{12}$$

where a bar denotes a pre-determined calibrated parameter. The GARCH(1,1) law of motion means that if $\bar{\psi} + \bar{\Upsilon} < 1$, then $E\left[\sigma_{it}^2\right] = \mu_i$, where the expectation is with respect to the stationary distribution. The hybrid approach deviates from standard GARCH by fixing the values of $\bar{\psi}$ and $\bar{\Upsilon}$ via calibration (calibration details are discussed in Appendix B.5). My focus allows 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. For estimation, the initial value, σ_1^2 must be fixed; a default in most statistical packages is to set $\sigma_1^2 = \mu$. Appendix A develops the asymptotic properties of this estimator, showing standard QML results apply.

The underlying standard set of parameters used to calibrate $\bar{\psi}$ and $\bar{\Upsilon}$ 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, I calculate such a standard set of parameters based on the monthly 128-variable McCracked & Ng FRED-MD database. The following simulation study shows that hybrid GARCH estimation performs well.

5 Performance of estimators

To evaluate the practical potential of TVV-ID, it is important to differentiate its performance from that of the 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. In doing so, I examine the performance of a variety of popular regime-estimation approaches in simulation; the results are highly varied. 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. In general, TVV-ID performs favorably, and, in particular, the likelihood-based estimators are superior.

5.1 The breakdown of conditional diagonality

When regime breaks must be estimated based on the data, estimates obtained via the Rigobon method face an important source of bias. It is standard in the SVAR context to assume that $E\left[\varepsilon_t \varepsilon_t'\right]$

Table 1: *The presence of off-diagonal elements*

	$E[\varepsilon_{1t}\varepsilon_{2t} \mid t \in T]$	$E[\varepsilon_{1t}\varepsilon_{2t} \mid t \in A]$	$E[\varepsilon_{1t}\varepsilon_{2t} \mid t \in B]$
$H = I_2$	-0.001	-0.002	0.000
$H = \left([1,1]', [1,1]'\right)$	-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 $\left([1,0]', [0,2]'\right)$ 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.

is diagonal; the Rigobon scheme requires in addition that $E[\varepsilon_t\varepsilon_t' \mid t \in A]$ is diagonal for a subsample, $A \subset T$, used for identification. 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, $E[\varepsilon_t\varepsilon_t' \mid t \in A]$ will not, in general, be diagonal. Put differently, the estimated regimes are not exogenous to ε_t as they are calculated based on η_t.

A simple example illustrates this fact. For this purpose, compare the generic $H = I_2$ to $H = \left([1,1]', [1,1]'\right)$. Let the "low variance" regime be I_2 and the high $\left([1,0]', [0,2]'\right)$; shocks are normally distributed. Now, compute the trace of $H\varepsilon_t\varepsilon_t'H'$ for each observation and define sub-sample A as those draws whose trace is above the overall median, and B below. I compute the expectations numerically using 500,000 draws for each regime. Table 1 reports the 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 orthogonality clearly breaks down, as described above.

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

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

$$= c_1 + H_{12}^2 c_2 + H_{12}c_3,$$

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, introducing additional confounding parameters. 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 $\Sigma_A \equiv$

$E[\Sigma_t \mid t \in A]$ 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.

This problem is likely to manifest if the true variance process exhibits continuous variation. In particular, estimated "high" regimes will mix individual short periods with high variances with surrounding periods of low variance, but with draws of ε_t conducive to large $\eta_t \eta_t'$; the converse is also true. The mixing of variance regimes is not a problem in itself (as shown in Rigobon (2003)), but the inclusion of periods on the basis of certain draws of ε_t is. This is potentially a problem in data with frequent variance changes, where the end observations of each regime, marginal based on variance value, are largely assigned based on ε_t . In smaller samples, these end observations may make a sizable contribution to the subsample expectations.

The researcher faces a trade-off in choosing regimes - either explicitly, or implicitly via the likelihood of a Markov switching model. 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 explores the bias of Rigobon identification based on rolling window variance 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. The model used for calibration is a bivariate SVAR(12) where the regressors are the first factor of the McCracken &

Table 2: *Calibration of volatility processes for simulations*

AR(1)

μ	ϕ	$E[e_t e_t']$	details
$(17.49, -42.11)'$	$(0.90, 0.96)'$	$\begin{bmatrix} 0.106 & 0.074 \\ 0.074 & 0.128 \end{bmatrix}$	$T = 100, 200, 400$ weak i.d.: $E[e_t e_t']/10$

GARCH(1,1)

μ	ψ	Υ	details
$(7.29, 0.46)'$	$(0.704, 0.764)'$	$(0.169, 0.226)'$	$T = 200$ weak i.d.: $\Upsilon/1.5$

Markov Switching

P_{trans}	V_{low}	V_{high}	details
$\begin{bmatrix} 0.971 & 0.029 \\ 0.092 & 0.908 \end{bmatrix}$	$\begin{bmatrix} 4.12 & 0 \\ 0 & 0.11 \end{bmatrix}$	$\begin{bmatrix} 15.91 & 0 \\ 0 & 2.27 \end{bmatrix}$	$T = 200$

Calibration of the volatility processes used in the simulation studies. Values of estimated using a bivariate VAR(12) based on the first factor of the McCracken & Ng FRED-MD database (excluding the FFR) and the FFR. H was first estimated using the QML AR(1) approach and the other structural parameters were estimated based on these structural shock series. The "weak" versions of the AR(1) and GARCH (1,1) are scaled to offer substantively less identifying variation, as discussed in Section C.3 of the Online Appendix.

Ng data (excluding the FFR), and the FFR. H is estimated using the QML AR(1) SV model, and given by

$$H = \begin{bmatrix} 1 & 0.298 \\ 0.333 & 1 \end{bmatrix}.$$

This H is then used to obtain time-series for ε_t from which the structural parameters are estimated for the other DGPs. Calibration details are in Table 2.[13] Shocks are normally distributed (as are innovations to the log SV process).[14]

The basic sample length is $T = 200$. This is shorter than the approximately 600 and 400 observations in the McCracken & Ng FRED-MD dataset and the empirical application respectively, because these sample sizes are somewhat long relative to many macroeconomic datasets. The shorter sample puts the methods through a sterner test in terms of strength of identification. Simulations are conducted with 5,000 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.

[13]For the calibration, the estimated GARCH and ARCH parameters were scaled by 0.99 for the second structural shock to avoid non-stationarity to machine precision.

[14]Unreported additional simulations show that using heavy-tailed shocks (implying misspecification for the likelihood estimators) does not significant harm estimates. In general, it actually improves performance for estimators that implicitly fit the variance of $\eta_t \eta_t'$, which now contains additional identifying information due to non-Gaussianity.

Study 1: Estimating regimes 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 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 volatility model with continuous variation. Estimation proceeds using the Sims methodology discussed in Section 5.1.

For both DGPs, many 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 to Monte Carlo error. Histograms for all estimators are available in Figures C.3-C.5 of the Online Appendix. Table 3 shows severe bias results when the regime determination is based on only the contemporaneous values of $\eta_t'\eta_t$, although RMSE is moderate. Note that without bias of the type discussed above, even these estimators should be consistent, like the oracle. Bias falls with the length of the rolling window, with lower RMSE for the 13-period window than 7-period. Performance for the SV DGP is similar; results are available in Table C.1 of the Online Appendix. Here, longer windows yield estimates quite close to the true values. That longer windows better identify the true parameters accords with theory; the off-diagonal bias will be minimized, since it is unlikely to have a long sequence of similarly conducive draws of ε_t driving the regime determination. Results appear most sensitive to window length, with the threshold and norm having smaller effects. However, these results suggest that even if the researcher has a strong belief that the underlying heteroskedasticity is dramatic (as it is here - see Section C.3 of the Online Appendix), care must be taken in estimating regimes, as there is potential for substantial bias.

Study 2: Comparison of estimators Three identification schemes are compared in this simulation study: the Rigobon approach, Sentana & Fiorentini's GARCH-based identification, and TVV-ID. Table 4 summarizes these estimators. The first implementation of Rigobon uses sub-samples defined based on a 13-period rolling window of $trace\left(\eta_t\eta_t'\right)$, with the regime cut-off being the median. This combination was chosen as it performed quite well in the Study 1. Second, it is implemented with the two sub-samples corresponding to simply the first and second halves of the sample. This should avoid the bias outlined above, but may be susceptible to weak identification. Third, I estimate a Markov switching model via maximum likelihood. The standard implementation of Sentana & Fiorentini's (2001) scheme uses maximum likelihood on GARCH(1,1) variance processes. TVV-ID is implemented in three ways. First, an approximation to quasi-maximum likelihood is estimated via Hamiltonian MCMC with flat priors and the model of equation (11), allowing for correlated innovations across dimensions. The point estimate is the median of the MCMC draws. GMM is

Table 3: *Mean estimates for Rigobon estimator: Markov-switching DGP*

	window	1-period			7-period			13-period			oracle	
norm	threshold	H_{21}	H_{12}	RMSE	H_{21}	H_{12}	RMSE	H_{21}	H_{12}	RMSE	H_{21}	H_{12}
trace	median	0.078	-0.24	3.79	0.016	0.419	4.93	0.023	0.357	4.42	0.033	0.277
	mean + 1 *s.d.*	0.064	-0.156	5.91	0.006	0.451	6.81	0.006	0.453	6.57		
$\bar{\eta_1^2}$	median	0.061	-0.032	1.91	0.009	0.441	6.10	0.018	0.392	5.22		
	mean + 1 *s.d.*	0.067	-0.139	4.20	0.009	0.427	7.12	0.011	0.403	6.83		

Mean estimates for Rigobon-type estimators for the empirically-calibrated Markov-switching DGP, $T = 200$, 5,000 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 estimate to the true H to minimize L_2 norm. Since the RMSE must account for error in multiple parameter estimates, the MSE is computed for each, and then normalized by the square of the true parameter, before the root of the sum is taken.

Table 4: *Estimators considered in simulations*

Identification scheme	Estimator	Summary
Rigobon	Sub-sample	13-month moving ave. trace, median threshold
	$T/2$ split	subsample split at $T/2$
	Markov switching	Maximum likelihood on Markov switching model
Sentana & Fiorentini	GARCH	Independent GARCH(1,1) processes, maximum likelihood
TVV-ID	Quasi-likelihood	AR(1) log SV with correlated innovations (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 estimators applied are split into three categories based on the identification approach exploited. The key features of each estimator are described in the summary column.

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 $\bar{\psi} = 0.6048, \bar{\Upsilon} = 0.2476$ (estimated based on 128 macro time series, see Appendix B.5), with the remaining estimation, including each process's mean, via QML.

Seven DGPs are considered. The three central specifications are a Markov switching process, a GARCH(1,1), and an AR(1) log SV, all with $T = 200$. I augment this with variants of the GARCH(1,1) and AR(1) SV. For GARCH(1,1), I add a version with "weak variation in the volatility; for the AR(1) SV, I consider $T = 100, 400$ and a version with $T = 200$ and "weak" variation in the volatility. The "weak" calibrations are chosen to scale down the ARCH parameters by a factor of 1.5 for GARCH(1,1) and the innovation variances by a factor of 10 for the AR(1) [15]. These values

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

are admittedly arbitrary, but were chosen after comparing sample variance paths under different calibrations to obtain paths that exhibited modest fluctuation in the neighborhood of the process mean. Representative sample paths are given in Section C.3 of the Online Appendix. 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 details are presented in Table 2.

The mean values obtained from each estimator, the RMSE, and rejection rates of the associated standard errors are reported in Table 5. Histograms for all DGPs are available in Figures C.6-C.12 of the Online Appendix. Estimates are in general better for H_{21} than H_{12}, which can be attributed to the larger autoregressive parameter $\phi_2 = 0.96$ for the second shock series. All estimators exhibit bias for certain DGPs (particularly in small samples), but for the most part, this is driven by a few outliers; the histograms demonstrate that the distributions are meaningfully centered around the true parameter values, with few exceptions. Given this, the RMSE is a useful tool in comparing the estimators, in giving a measure of their dispersion around the true parameter values.

Across DGPs, the QML implementation of the AR(1) SV model performs best. It does exhibit some bias due to outliers for smaller sample sizes, but even when misspecified, its RMSE is only slightly worse than those of the correctly specified estimators. The once case where it breaks down, with substantial bias, is for H_{12} in the weak GARCH DGP, where it is both misspecified and faced with weak identification, but there it still maintains an RMSE not too far behind the leading well-specified GARCH estimator. Its robustness to misspecification - that, even when misspecified - it can compete with well-specified estimators, is not shared by any other estimator. The rejection rates are actually undersized, offering conservative inference.[16]

The hybrid GARCH estimator and Markov switching estimators offer the next best performance. Frequently, they actually have a lower bias than the AR(1) SV due to fewer outliers, but their RMSEs are systematically higher. Their performance varies with the degree of misspecification and sample size, as expected, and both break down in the presence of weak identification. The hybrid GARCH has some difficulty in fitting the "weak" GARCH DGP as it is calibrated to have a very different ARCH parameter. The standard errors for both estimators offer minimal size distortions.

The GARCH estimator generally is comparable to the hybrid GARCH and Markov switching approaches. However, it breaks down badly for the AR(1) with $T = 400$. This is because the empirical calibration dictates GARCH parameters that are very close to non-stationarity. As a result, with a longer draw of data, the dynamics sometimes appear explosive from a GARCH-fitting perspective, negatively impacting the estimates.[17] As opposed to being an artifact of the calibration, this should be seen as a strike against GARCH since these results follow from an empirical calibration

[16] One exception is for H_{12} in the AR(1) 400 DGP. This is likely because with a longer sample size the sampling variation is quite low, while given the unsupervised nature of the simulation study, some of the chains may not converge, or cycle between column orderings, inflating the rejection rate. This will not be a problem when MCMC is supervised as in usual practice.

[17] The start values for the GARCH parameters in the optimization are based on the median of 1000 sample paths of T drawn from the DGP. These start values, some of which are very large for $T = 400$, thus only impact the optimization commensurately with how they impact the data being passed to the algorithm.

Table 5: *Mean estimates and rejection rates: study 2*

		QL AR(1)		GMM		Hybrid		GARCH		Sub-sample (rolling)		Sub-sample ($T/2$)		Markov switching	
		mean	α	mean	α	mean	α	mean	α	mean	α	mean	α	mean	α
Markov switching, $T = 200$	H_{21}	0.038	1.2	0.018	33.1	0.024	9.6	0.025	48.5	0.023	0.0	0.014	0.1	0.034	4.4
	H_{12}	0.327	13.1	0.352	36.5	0.325	11.3	0.346	46.8	0.357	0.0	0.378	0.2	0.273	5.0
	RMSE	2.39	-	6.36	-	5.34	-	4.74	-	4.42	-	6.62	-	2.19	-
GARCH(1,1), $T = 200$,	H_{21}	0.037	1.1	0.030	13.4	0.031	4.3	0.033	4.7	0.028	0.0	0.029	3.4	0.031	11.6
	H_{12}	0.394	5.2	0.335	18.7	0.329	5.3	0.295	4.6	0.304	0.0	0.360	3.2	0.331	11.2
	RMSE	2.75	-	5.82	-	2.47	-	2.58	-	6.62	-	6.88	-	5.31	-
GARCH(1,1), $T = 200$, weak	H_{21}	0.041	0.5	0.024	49.8	0.023	24.5	0.032	4.9	0.033	0.0	0.022	0.1	0.026	9.5
	H_{12}	1.186	15.8	0.775	51.8	0.979	24.4	0.266	5.7	0.114	0.0	0.829	0.2	0.635	8.8
	RMSE	7.26	-	12.13	-	8.27	-	6.92	-	7.99	-	13.3	-	11.15	-
AR(1), $T = 100$	H_{21}	0.048	0.8	0.017	42.9	0.027	9.2	0.026	22.4	0.020	0.0	0.023	0.5	0.023	11.7
	H_{12}	0.441	2.2	0.585	46.0	0.347	10.2	0.364	21.5	0.445	0.0	0.387	0.3	0.438	10.4
	RMSE	4.30	-	8.92	-	6.67	-	6.37	-	8.28	-	7.42	-	7.26	-
AR(1), $T = 200$	H_{21}	0.041	0.9	0.023	37.2	0.030	6.9	0.031	23.6	0.021	0.0	0.024	0.4	0.030	5.8
	H_{12}	0.370	3.0	0.376	40.1	0.322	7.7	0.306	22.3	0.369	0.0	0.365	0.2	0.306	5.1
	RMSE	2.54	-	7.14	-	4.31	-	3.92	-	6.91	-	6.30	-	5.16	-
AR(1), $T = 400$	H_{21}	0.034	1.3	0.028	27.8	0.033	4.3	0.061	62.2	0.023	0.0	0.024	0.3	0.0343	4.7
	H_{12}	0.300	11.4	0.314	31.4	0.287	4.9	0.756	51.7	0.308	0.0	0.369	0.3	0.285	4.9
	RMSE	1.34	-	5.60	-	2.52	-	8.50	-	6.23	-	5.72	-	3.25	-
AR(1), $T = 200$, weak	H_{21}	0.063	0.2	0.014	42.7	0.0187	47.6	0.019	8.6	0.032	0.0	0.018	0.0	0.019	15.4
	H_{12}	0.44	3.2	0.535	43.0	0.503	48.6	0.500	10.0	0.155	0.0	0.470	0.0	0.465	14.3
	RMSE	4.85	-	8.96	-	8.36	-	7.78	-	7.17	-	8.41	-	8.04	-

Mean 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 presented for a nominally-sized 5% test 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 wild bootstrap. For GARCH, the Fisher information is used. For the hybrid, the QML variance is used. For Markov Switching, a parametric bootstrap is used. Since the RMSE must account for error in multiple parameter estimates, the MSE is computed for each, and then normalized by the square of the true parameter, before the root of the sum is taken.

based on commonly-used data. In some cases, estimates of the GARCH and ARCH parameters on the explosive edge of the parameter space lead to excess mass of H around zero. These are examples of problematic misspecification resulting from assuming GARCH as in Sentana & Fiorentini's approach; GARCH(1,1) struggles to accommodate a stationary DGP ($\phi_2 = 0.96$ is quite reasonable) of a different functional form. The rejection rates, based on the Fisher Information, are accurate when well-specified, but as expected, break down when misspecified.

GMM has difficulty with small sample sizes and weak identification. This makes sense, as there are fewer explicitly-estimated moments compared to the implicitly incorporated moments of the likelihood-based approaches. Thus, when there is difficulty precisely estimating and decomposing the higher moments of the shocks, identification is impacted. For AR(1) SV $T = 200$ though, the performance is very good. The standard errors also improve with sample size as the asymptotics "kick in". For larger sample sizes, GMM can offer a real alternative that requires no parametric assumptions.

The rolling window Rigobon estimator is generally reliable, which, given that the window length and threshold were chosen based on Study 1 to perform well in this data is unsurprising. A naïve implementation would perform very differently. When identification is strong, the bias is actually very low. However, the breakdown is dramatic for weak identification, which makes sense, as in this context the realized shock values dominate the regime selection, causing non-diagonality bias. Even when well specified (the Markov switching DGP), it is not competitive with the other estimators in terms of RMSE; the RMSE is high in general. The same remarks apply to the $T/2$ split estimator. The wild bootstrap used for both is severely undersized.

6 Empirical illustration: estimating oil price shocks

Kilian (2009) contributes to a substantial literature examining the effects of oil price fluctuations on the macroeconomy. In particular, Kilian (2009) examines how the effects may vary depending on the cause of the price change. He estimates a 3-variable SVAR (oil production growth, real economic activity, real oil price) to obtain structural shock series, each of which represents a possible cause of oil price fluctuations. Based on his structural assumptions, these are an oil supply shock, an aggregate demand shock, and an oil-specific demand shock. Examining the impact of these structural shocks on GDP and inflation illustrates that oil price movements have different effects on the real economy depending on the source of the price movement - which type of shock caused it.

This problem is a natural candidate for identification via time-varying volatility. It has in fact already been evaluated in a similar context by Lütkepohl & Netšunajev (2014) who assess the robustness of results using identification via heteroskedasticity with a Markov switching implementation. Moreover, as discussed in that paper, and in Zhu (2017), there is evidence that oil markets exhibit time-varying volatility. Further, the recursive identification assumptions previously exploited, while plausible, are not ex ante obviously true in a "hard zero" sense. TVV-ID allows me to finally perform such tests of identifying assumptions and more broadly to assess the robustness of results (in terms of

ultimate impulse responses) to a structural response matrix that is not lower triangular. The ability to test conventional identification assumptions on the H matrix is a key avenue for application of TVV-ID in applied work.

Explicitly, Kilian's (2009) model takes the form

$$\eta_t = \begin{pmatrix} \eta_t^{\Delta prod.} \\ \eta_t^{r.e.a.} \\ \eta_t^{r.p.o.} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ H_{21} & 1 & 0 \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_t^{oil\ supply} \\ \varepsilon_t^{agg.\ demand} \\ \varepsilon_t^{oil\ spec.\ demand} \end{pmatrix},$$

where the residual η_t is obtained from a 24-lag VAR of the percent change in global crude oil production, an index of real economic activity, and the real price of oil. Note that I have used a different normalization (unit diagonal of H as opposed to identity structural shock covariance) to remain consistent with the theoretical portion of the present paper, and because an identity covariance normalization (of a specific period) is not useful in the TVVID argument based on autocovariance. The data is monthly, spanning 1973:2-2007:12. The identifying assumptions are

1. Oil production does not respond to either aggregate or oil demand shocks within the month
2. Real economic activity does not respond to oil demand shocks within the month

These assumptions have also been adopted in Kilian & Park (2009).

I adopt Kilian's (2009) VAR specification, using replication code available from the AEA website, up to the estimation of the reduced form residuals. I deviate in terms of assuming time variation in the structural shock variances and estimating the H matrix using several forms of TVV-ID, instead of simply taking a Cholesky decomposition. Note that for identification purposes, I need not take a stand on whether the time variation is conditional or unconditional. I consider both the AR(1) stochastic volatility implementation, and the fully non-parametric 2-step GMM approach. Recall from the simulation study that the AR(1) stochastic volatility implementation, while an unconditional heteroskedasticity model, performs fairly well even in the presence of conditional heteroskedasticity, and that the GMM implementation is completely agnostic to the type of heteroskedasticity. For contrast, I also present results from a Rigobon (2003) estimator using the same 13-period rolling window and threshold rule as in the simulation study. Lütkepohl & Netšunajev (2014) conduct essentially the same exercise with a Markov switching model, so I do not pursue that here.[18] Table 6 presents the H matrices computed using each of these three approaches, with Kilian's (2009) result for comparison. Note that I employ the unit-diagonal normalization, while Kilian (2009) uses the unit-variance normalization.

Columns are labeled to provide the closest match to Kilian's (2009) results. Equivalently, labeling based on ruling out implausible magnitudes of responses leads to the same permutations of columns, due to the large differences in relative magnitude of the entries in each column. At first glance, it

[18] They do, however, consider a VAR(3) as opposed to the VAR(24), referring to several information criteria supporting a more parsimonious specification.

Table 6: *Estimates of H in three-variable oil market SVAR*

	AR(1) SV			GMM			Markov switching			Kilian		
	1	0.37	0.02	1	-0.14	-0.36	1	-0.09	0.42	1	0	0
H	-0.01	1	0.03	0.01	1	0.03	-0.00	1	0.14	0.00	1	0
	-0.01	0.04	1	0.02	0.09	1	-0.06	-0.20	1	-0.02	0.12	1
null	Chol.	col. 2	col. 3	Chol.	col. 2	col. 3	Chol.	col. 2	col. 3	_	_	_
$p-$val.	0.31	0.72	0.21	0.94	0.97	0.83	0.95	0.90	0.85			

The first block presents estimates for the H matrix based on the various estimators. The lower block reports p−values for tests of Cholesky structure collectively, and then by column. The AR(1) SV method implements an AR(1) model for log-variance using MCMC. Standard errors are based on the sampling variation of the chains. The GMM approach uses 2-stage GMM. The Rigobon method employs a 13-month rolling window with the median trace as the regime threshold; standard errors use a wild bootstrap. The Kilian matrix is a Cholesky decomposition, with unit normalized diagonal.

appears that the matrices obtained without restrictions differ from Kilian's (2009) due to non-zero point estimates in the upper-triangular portion.

However, I can now test the recursive assumptions that Kilian (2009) uses to obtain structural responses, as they are overidentifying restrictions when the model is identified using TVVID. Proposition 3 presents the test formally.

Proposition 3. *Under Assumptions 1.1-1.2, 2′, & 3,*

1. *Conventional identifying assumptions on the H matrix of the form $\Lambda H = \lambda$ (where Λ is $p \times n$ and λ $p \times 1$) are overidentifying restrictions,*

2. *For an asymptotically normally distributed estimator $\hat{H}$ with mean H and asymptotic variance var $\left(\hat{H}\right)$, these overidentifying restrictions can be tested using the Wald test statistic*

$$J_\Lambda = \left(\Lambda\hat{H} - \lambda\right)' \left[\Lambda var\left(\hat{H}\right)\Lambda'\right]^{-1} \left(\Lambda\hat{H} - \lambda\right),$$

where $J_\Lambda \xrightarrow{d} \chi^2_p$ under the null hypothesis $\Lambda H = \lambda$.

The first point follows from Theorem 1. The test statistic and its limiting distribution follow from the properties of a standard Wald test. Accordingly, I perform a simple joint Wald test for the three overidentifying zero restrictions, and report separate test results for the restrictions on each column of H. The asymptotic variance for the AR(1) SV estimator is based on the sampling variation from the MCMC and for the Rigobon estimator, a recursive wild bootstrap.[19] For both AR(1) SV and GMM, the recursive assumptions cannot be rejected column-by-column or collectively.

[19]Appendix A discusses the use of QML-type standard errors for MCMC as proposed by Müller (2013). Note that the simple standard errors based on the sampling distribution were found to be conservative in the simulation study. The Müller (2013) errors were also computed here for robustness and found to be even more conservative than those reported, so do not change the conclusions.

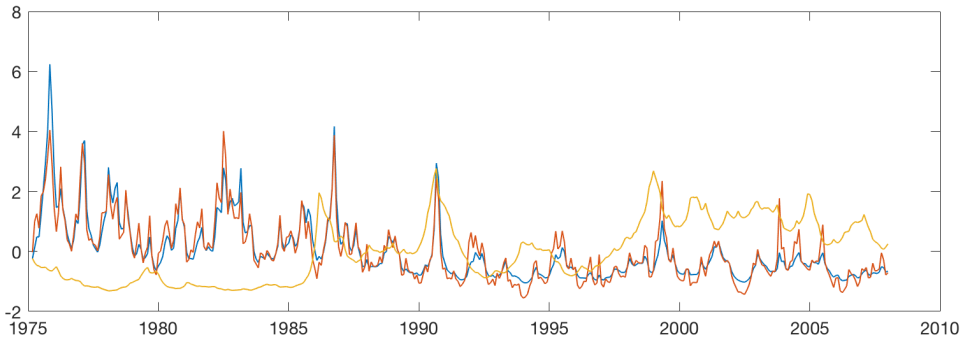

Figure 2: *Estimated variance paths for the three structural shocks obtained from the AR(1) SV method, normalized to present on the same axes. The blue line represents the variance of the oil supply shock, red aggregate demand, and gold oil-specific demand.*

This should be particularly convincing in the face of the size distortions found in the simulation study for the GMM standard errors in samples of this size. The results using the Rigobon estimator also cannot reject the Cholesky structure, though this is unsurprising given how undersized the bootstrap approach was found to be in simulation. Thus, I cannot reject the validity of Kilian's (2009) lower triangular structure. These results accord with those of Lütkepohl & Netšunajev (2014), based on identification via heteroskedasticity using a Markov switching model; they are also unable to reject the overidentifying restrictions of a lower triangular structure. Comparing the implied paths of the structural shocks shows very similar results across both identification schemes and estimators.

The AR(1) SV implementation allows me to recover implied paths for the shock variances. Figure 2 plots these paths, standardized to display on the same axes. The oil supply shock variance is in blue, the aggregate demand in red, and the oil-specific demand in gold. The general dynamics of these accord with the labeling of the shocks and economic intuition. Oil supply shocks are at their most volatile during the first part of the sample, and become somewhat more stable in the latter period. Aggregate demand shock volatility behaves quite similarly to that of oil supply shocks.[20] Oil demand shock volatility rises throughout the sample, punctuated by periodic spikes. For a comprehensive discussion of events affecting oil markets during this period, see e.g. Barsky & Kilian (2002, 2004)

Besides testing the impact of the recursive ordering assumptions on the H matrix itself, it is important to assess their impact on the impulse response functions. Figure 3 reports the IRFs under the four identification approaches. Kilian's (2009) responses are in gold, AR(1) SV in blue, GMM in red, and Rigobon in purple. Generally speaking, the paths are all very similar; this should not be surprising, considering the role of reduced form parameters in dictating the IRFs. The largest deviations occur for the Rigobon estimates; potential issues with this methodology have already been discussed at length. Figure 3 replicates Kilian's (2009) Figure 3 under the AR(1) SV

[20]Note that the similar paths are not a major issue for identification since identification is based on a combination of both autocovariance, mean and other implied moments in this implementation, which do not collectively exhibit collinearity. On the other hand, such data would challenge Sentana & Fiorentini's (2001) identification approach, which requires linear independence of the variance paths.

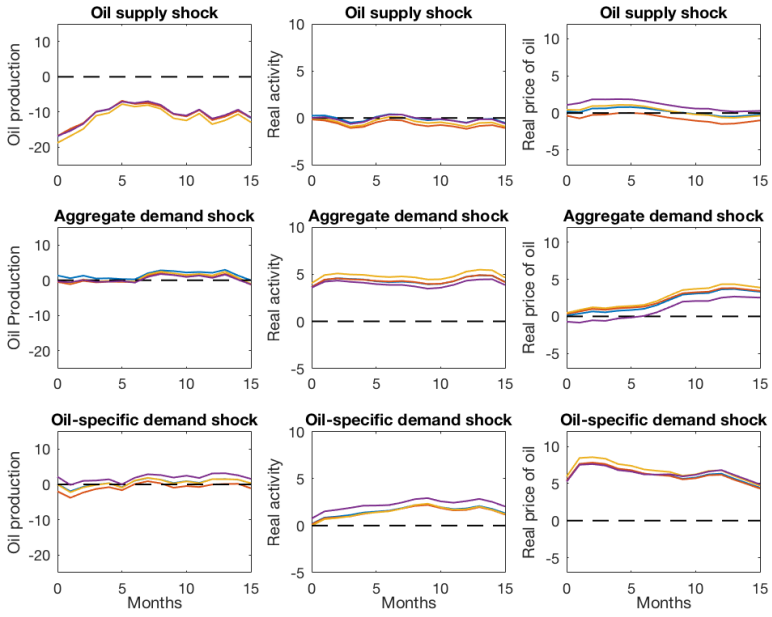

Figure 3: *Comparison of IRFs to oil price shocks for each of the four estimators, replicating the IRFs from Kilian (2009) Figure 3. The blue lines correspond to AR(1) SV, red to GMM, purple to Rigobon, and gold to Kilian. Shocks are one standard deviation, which is calculated separately for each path based on the structural shocks implied from the respective unit-diagonal H matrices.*

implementation of TVV-ID (blue), with Kilian's results for comparison (gold, dashed), and 1- and 2-standard deviation confidence intervals. The procedure for the confidence intervals is discussed in Appendix B.4. The only possible case for rejecting Kilian's (2009) responses is at early horizons, when the mild bias caused by imposing recursive assumptions is at its highest. This is also impacted in level shifts by the fact that a "one standard deviation shock" has a different size in each of the two implementations.

Kilian's (2009) analysis culminates with the response of GDP and inflation to each type of oil shock, making the argument that the macroeconomic impact of oil price fluctuations depends on the type of shock causing them. I replicate these responses (Kilian (2009) Figure 5) under TVV-ID in Figure 5, with TVV-ID in blue and Kilian (2009) in gold, dashed. Like the impulse responses of the three series internal to the SVAR, these are again very similar to those in the original paper. The economic conclusions remain unchanged: oil supply disruptions may have a small negative impact on GDP, but little impact on inflation; aggregate demand shocks have a small stimulatory effect at first, before higher prices have a recessionary impact; oil demand shocks orthogonal to macroeconomic fundamentals are contractionary and inflationary. This empirical illustration validates and indeed strengthens the results of Kilian (2009) by establishing that they are robust to alternative identification assumptions, in particular relaxing sharp recursive assumptions on the H matrix, and replacing them with *estimated* zeros. However, the fact that estimates of the H matrix can be obtained that match existing results (with structural assumptions used only to interpret the shocks) should not be discounted. The potential to test existing identification

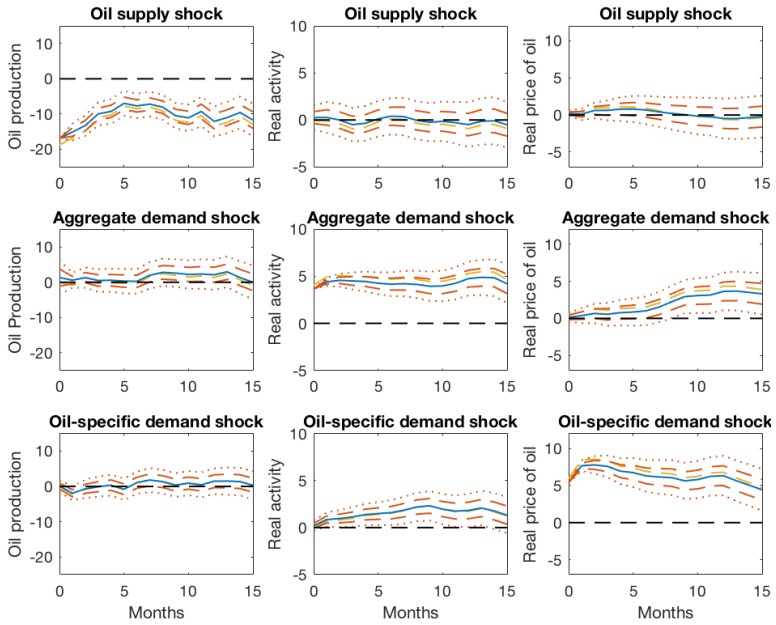

Figure 4: *IRFs are constructed using the estimates based on the AR(1) SV process, replicating the IRFs from Kilian (2009) Figure 3. 1- and 2- standard deviation confidence intervals are also plotted. The point estimate is in blue, the confidence intervals in red, and the Kilian (2009) path in gold, dashed. The standard errors combine Kilian's (2009) wild bootstrap for the reduced form IRFs with the covariance of the H estimates (based on sampling distribution), as described in Appendix B.4.*

assumptions as overidentifying restrictions is valuable. All identification assumptions are not created equal, and in other contexts the Cholesky structure is likely to be rejected with the help of TVV-ID.

7 Conclusion

This paper develops a general framework under which latent shocks can be identified via time-varying volatility. The previous literature offers identification arguments based only on parametric models of 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 illustration based on Kilian's (2009) study of the effects of oil price-driving shocks on the macroeconomy illustrates a key usage of TVV-ID. By estimating the H matrix, relying only on the presence of time-varying volatility, I am able to directly test the Cholesky structure frequently deployed in this literature. I show that TVV-ID produces very similar results to those in

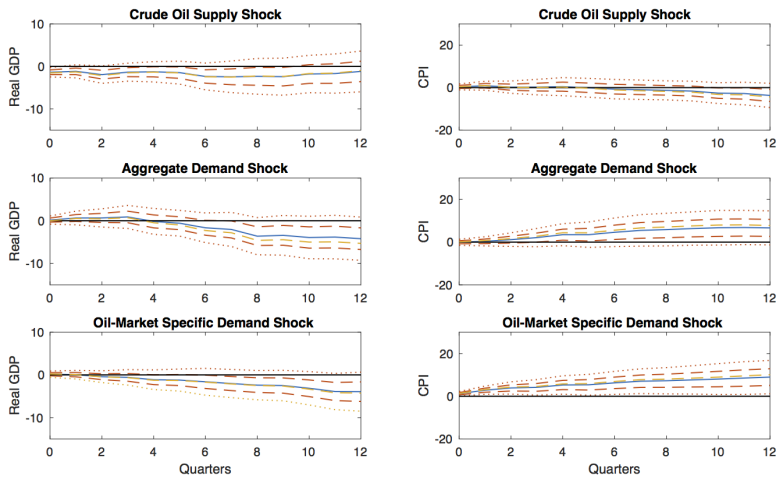

Figure 5: *The IRFs plot the impact of the three oil price shocks on US real GDP and inflation at quarterly horizons, replicating the IRFs from Kilian (2009) Figure 5. Estimates are based on the AR(1) SV estimator, with 1- and 2- standard deviations confidence intervals. The point estimate is in blue, the confidence intervals in red, and the Kilian (2009) path in gold, dashed. The standard errors follow Kilian's (2009) block bootstrap.*

the original paper, and tightly estimates zeros instead of assuming them. In this case, I am able to validate Kilian's (2009) results by demonstrating their robustness to identifying assumptions. However, in other contexts, this will not be the case, and TVV-ID can help to put into sharp relief those assumptions on the H matrix that ought be deemed unreliable.

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.

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.

Barsky, R. and L. Kilian (2002): "Do we really know that oil caused the great stagflation? A monetary alternative," in *NBER Macroeconomics Annual 2001, Volume 16*, Chicago: University of Chicago Press, 734, 137–183.

Barsky, R. B. and L. Kilian (2004): "Oil and the Macroeconomy Since the 1970s," *Journal of Economic Perspectives*, 18, 115–134.

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

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.

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.

Brown, R. G. and P. Y. Hwang (1996): *Introduction to Random Signals and Applied Kalman Filtering*, New York: John Wiley and Sons, 3 ed.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2017): "Feedbacks: Financial Markets and Economic Activity," manuscript, 1–50.

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

Carriero, A., T. E. Clark, and M. Marcellino (2017): "Endogenous Uncertainty?" manuscript.

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.

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): "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," *The Journal of Economic Perspectives*, 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.

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.

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

Hamilton, J. D. (2003): "What is an oil shock?" *Journal of Econometrics*, 113, 363–398.

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)*, 107, 3119–3145.

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.

——— (2008a): "A Comparison of the Effects of Exogenous Oil Supply Shocks on Output and Inflation in the G7 Countries," *Journal of the European Economic Association*, 6, 78–121.

——— (2008b): "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics*, 90, 216–240.

——— (2008c): "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature*, 46, 871–909.

——— (2009): "Not All Oil Price Shocks Are Alike : Disentangling Supply Shocks in the Crude Oil Market," *The American Economic Review*, 99, 1053–1069.

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, Cambridge: Cambridge University Press.

Kilian, L. and C. Park (2009): "The Impact of Oil Price Shocks on the U.S. Stock Market," *International Economic Review*, 50, 1267–1287.

Kilian, L. and C. Vega (2011): "Do Energy Prices Respond to U.S. Macroeconomic News? A Test of the Hypothesis of Predetermined Energy Prices," *Review of Economics and Statistics*, 93, 660–671.

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.

——— (2010): "Structural Vector Autoregressions with Nonnormal Residuals," *Journal of Business and Economic Statistics*, 28, 159–168.

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.

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 (2018): "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.

Lütkepohl, H. and A. Netšunajev (2017): "Structural Vector Autoregressions with Smooth Transition in Variances: The interaction between US monetary policy and the stock market," *Journal of Economic Dynamics and Control*, 84, 43–57.

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

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.

Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2018): "Inference in Structural VARs with External Instruments," *Quarterly Journal of Economics*, forthcomin.

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 (2018): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics (forthcoming)*.

Netšunajev, A. and H. Lütkepohl (2014): "Disentangling Demand and Supply Shocks in the Crude Oil Market: How to Check Sign Restrictions in Structural VARs," *Journal of Applied Econometrics*, 29, 479–496.

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

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.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions," *Review of Economic Studies*, 72, 821–852.

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.

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.

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

Stock, J. H. and M. W. Watson (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.

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.

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.

Zhu, B. (2017): "Forecasting the real price of oil under alternative specifications of constant and time-varying volatility," *CAMA Working Papers*, 71.

Appendix

A Proofs

Derivation of the simple Rigobon estimator

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

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

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

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

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:

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

Derivation of Proposition 1

Proof. I start with

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

Since v_t was shown to be a martingale difference sequence and $\mathrm{Var}_t\left(v_t\right) < \infty$ (Assumption 3.2),

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

This also implies that in the signal-noise decomposition, Equation (6), v_t is white noise. Using this fact, Assumption 2, Assumptions 3.1-2, and the decomposition of ζ_t above, it is immediate that, for $s \neq t$,

$$\begin{aligned}E_{t,s}\left(\zeta_t\zeta_s'\right) = {} & L\left(H \otimes H\right) GE_{t,s}\left[\sigma_t^2\sigma_s^{2'}\right] G'\left(H \otimes H\right)' L' \\ & + L\left(H \otimes H\right) GE_{t,s}\left[\sigma_t^2 v_s'\right] + E_{t,s}\left[v_t\sigma_s^{2'}\right] G'\left(H \otimes H\right)' L'. \end{aligned} \tag{13}$$

By the law of iterated expectations, Assumption 1.1 implies that

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

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

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

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

$$E_{t,s}\left(\zeta_t\zeta_s'\right) = L\left(H \otimes H\right) GE_{t,s}\left[\sigma_t^2\sigma_s^{2'}\right] G'\left(H \otimes H\right)' L' + L\left(H \otimes H\right) GE_{t,s}\left[\sigma_t^2 v_s'\right]. \tag{14}$$

Finally, I can rewrite (14) as

$$L\left(H\otimes H\right)\left(GE_{t,s}\left[\sigma_t^2\sigma_s^{2'}\right]G'+GE_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)\right]\right)\left(H\otimes H\right)'L'$$

$$=L\left(H\otimes H\right)G\tilde{M}_{t,s}\left(H\otimes H\right)'L'$$

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

Derivation of Proposition 2

Proof. It is necessary to show that given Assumption 4, $E_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)'\right]=E_{t,s}\left[\sigma_t^2 u_s'\right]$ $\times G'$ where $u_s=matdiag\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)$. Assumption 4 states that $E_{t,s}\left[\sigma_{it}^2\left(\varepsilon_s\varepsilon_s^{'}-\Sigma_s\right)\right]$ is diagonal for all $i=1,2,\ldots,n$. Therefore, $E_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)'\right]$ 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 $E_{t,s}\left[\sigma_t^2\left(\varepsilon_{js}^2-\sigma_{js}^2\right)\right]$. Now by the definition of G, AG' takes the j^{th} 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 j^{th} column of A is equal to $E_{t,s}\left[\sigma_t^2\left(\varepsilon_{js}^2-\sigma_{js}^2\right)\right]$ for all $j=1,2,\ldots,n$, then $AG'=E_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)'\right]$. This is true if $A=E_{t,s}\left[\sigma_t^2 diag\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)\right]=E_{t,s}\left[\sigma_t^2 u_s'\right]$. Thus $E_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s\varepsilon_s'-\Sigma_s\right)'\right]=E_{t,s}\left[\sigma_t^2 u_s'\right]G'$ as desired. Proposition 2 follows from re-writing the two terms of (13) in this way, and summing to yield $\check{M}_{t,s}$. □

Proof of Theorem 1

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

Definition 1. *Define*

1. $U_1D_UU_2'=V$, *a reduced SVD,* V $n_1\times n_2$, D_U $d\times d$,
2. C_i *is a full rank matrix,* $m_i\times n_i, m_i\geq n_i$,
3. $F=C_1VC_2'$, *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 $Q_1R_1=CU_1$, a reduced QR decomposition, and similarly for CU_2. Then $F=Q_1R_1D_UR_2'Q_2'$. 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 $W_1D_RW_2'=R_1D_UR_2'$, another SVD; then $F=\left(Q_1W_1\right)D_R\left(W_2'Q_2'\right)$ is a reduced SVD (it is easily shown D_R are singular values of F, and the corresponding vectors are clearly orthogonal). Thus write $Q_1R_1\left(R_1^{-1}W_1\right)=Q_1W_1$ 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*

1. *In particular, $C_1 = (H \otimes H) G$, $n^2 \times n$ with rank n,*
2. *G is a selection matrix such that $vec(ADA') = (A \otimes A) G\, diag(D)$,*
3. *$\hat{S}_1 = C_1 U_1 \Gamma_1 T_1$, an arbitrary reduced SVD of F,*
4. *V is $n \times n^2$ and has no scalar multiple rows,*
5. *$rank(V) \geq 2$.*

Proposition 4. 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 j^{th} 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 4 is re-written in economic terms to yield Theorem 1.

Proof of Corollary 1

Proof. Corollary 1 follows directly from Proposition 4 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 $i = 1, 2, ..., d$. □

Proof of Theorem 2

Proof. Theorem 2 is based on the argument underlying Proposition 4. Note that the vectorization of $E_t[\zeta_t]$ is given by $vech(HE_t[\Sigma_t]H')$, an additional equation of the form found in CU_1. Define $U_{1,M}D_MU'_{2,M} = M_{t,s}$ 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. □

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 $g\left(H^{-1}\eta_t\right)$. 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

$$Cov_{t,s}\left(\sigma_t^2, \sigma_s^2\right) = E_s\left[vec\left(H^{-1}\eta_t\eta_t'\left(H'\right)^{-1}\right)vec\left(H^{-1}\eta_s\eta_s'\left(H'\right)^{-1}\right)'\right].$$

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.

Asymptotic properties of the hybrid GARCH estimator

The form of the hybrid GARCH estimator is described in (12). I now establish the asymptotic properties of this estimator. Define the filtration $\mathcal{F}_{t-1} = \left\{\sigma_1^2, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{t-1}\right\}$, $\mathcal{F}_0 = \sigma_1^2 = \mu$. Stacking (12) gives

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

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

$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\left(0, H\Sigma_t\left(\mu, \mathcal{F}_{t-1}\right) H'\right), \, t = 1, 2, \ldots T, \tag{15}$$

recalling $\eta_t = H\varepsilon_t$ and $\Sigma_t\left(\mu, \mathcal{F}_{t-1}\right) = diag\left(\sigma_t^2\left(\mu, \mathcal{F}_{t-1}\right)\right)$. It is straightforward to maximize the joint quasi-likelihood for $t = 1, 2, \ldots, T$ with respect to μ 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 μ, 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 $g\left(\eta_t \mid \mathcal{F}_{t-1}; \theta\right)$. Denote the model joint density, the multivariate normal density implied by (15), as $f\left(\eta_t \mid \mathcal{F}_{t-1}; \theta\right)$. Define $L_T\left(\theta\right) = \frac{1}{T}\sum_{t=1}^{T} E_G\left[\log f\left(\eta_t \mid \mathcal{F}_{t-1}; \theta\right)\right]$, where $E_G\left[\cdot\right]$ denotes the unconditional expectation with respect to G. It is well known that maximizing $L_T\left(\theta\right)$ with respect to θ is equivalent to minimizing the Kullback-Leibler distance with respect to θ. Denote θ^* as the unique maximizer of $L_T\left(\theta\right)$. The sample counterpart of $L_T\left(\theta\right)$ is $\bar{L}_T\left(\eta_t; \theta\right) = \frac{1}{T}\sum_{t=1}^{T} \log f_t\left(\eta_t \mid \mathcal{F}_{t-1}; \theta\right)$ 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 5.

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

Assumption 5.3-4 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, $E_G\left|f_t\left(\eta_t \mid \mathcal{F}_{t-1}; \theta\right)\right| < \infty$ for all $t = 2, 3, \ldots, T$ since $\sup_{\eta_t \mid \eta_{t-1}}\left|f\left(\eta_t \mid \mathcal{F}_{t-1}; \theta\right)\right| < \infty$. Together with Assumption 5.2, 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 6.

1. *θ^* is a regular point of $D(\theta) = E_G\left[\nabla^2 \log f(\eta_t \mid \mathcal{F}_{t-1}; \theta)\right]$,*
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 Assumption 5.2-3) 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 6, White (1982) Theorem 3.2 gives

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

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

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, $C(\theta_0) = -D(\theta_0)^{-1}$.

B Considerations for application

B.1 Labeling

In the text, I provide a sketch of possible labeling assumptions for the structural shocks, or, equivalently, identified columns of the H matrix. I develop them in more detail here, with examples. 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. 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 $E[Z_t \varepsilon_{jt}] = 0$ for $j \neq i$, and non-zero for $j = i$, the i^{th} 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 $\frac{var(\hat{Z}_t^i)}{var(Z_t)} > \frac{var(\hat{Z}_t^j)}{var(Z_t)}$ 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 flavor 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. 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 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."

B.2 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.3, 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.3 Practical points on estimation methods

B.3.1 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).[21]

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's Hamiltonian MCMC 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 different parameter values that are observationally equivalent, but correspond to different labelings of the columns of H, or alternate between labelings during the chain. It is thus essential to inspect the chains and label the shock series estimated by each chain separately before combining the chains to compute overall estimates, to avoid averaging over different labelings of the columns of H. Some experimentation is necessary, depending on the dimension and distribution of data, to determine how many iterations are needed for both warmup and convergence to the stationary distribution. The main results in my empirical application are based on five chains of 10,000 iterations, each with a 2500-iteration warm-up. If the researcher has priors over the model parameters, these can be incorporated to perform fully 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.[22]

These methods have the advantage of directly computing a distribution of values for the filtered volatility path, which is a potentially interesting object in itself. 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

[21] 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.

[22] 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.

policy-making, in a way that the previous custom of simply looking at implied realized 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, I unaware of any 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).

Müller (2013) offers an analog to 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 can be estimated with the sampling covariance from the MCMC. The score can be evaluated based on 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 can then be estimated using a HAC estimator, like the equal-weighted-periodogram advocated in Chapter **??**. Standard errors following this approach were constructed but not reported for the empirical illustration; naturally, they are more conservative, and those computed simply from the sampling distribution of the MCMC procedure already fail to reject the Cholesky structural assumptions.

B.3.2 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^2_t; the matrix $M_{t,s}$ is simply 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 information alone could not identify H.[23] Also recall that additional moments reduce the possibility of weak or non-identification, as discussed in Theorem 2. For standard GMM asymptotic results to apply, the σ^2_t 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

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

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. Nevertheless, the simulation study in Section 5 shows that GMM can perform quite well for larger sample sizes. Given the appeal of this completely non-parametric approach, it should be considered in such settings.

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.3.3 GARCH

With my new results demonstrating identification for any functional form (that implies an auto-covariance), it is worth reconsidering the role GARCH can play. Since identification is no longer reliant on GARCH assumptions in the knife-edge sense of Sentana & Fiorentini (2001), 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 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 some settings. Primarily, this is due to excess mass around zero in the distribution of the estimator and difficulty handling (stationary) DGPs whose GARCH approximation may appear non-stationary. 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.4 Standard errors

Each of the estimation approaches 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.

For SVARs, it is a familiar result that the asymptotic covariance matrix of moments for the autoregressive coefficients and the covariance decomposition (the estimation of the H matrix) 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 employed here. First, consider GMM (with the assumption of fourth-order stationarity of σ_t for expositional simplicity). The off-diagonal blocks take the form

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

where the last equality follows from the fact that η_u is uncorrelated across time, and thus, asymptotically, has zero covariance with the autocovariance of $\eta_t\eta_t'$. 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, provided η_t is symmetrically distributed. 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 so η_t is symmetrically distributed, then the result follows from $E\left[\eta_t \mid \eta_s\eta_s', Y_{t-j}\right] = 0, j = 1, 2, \ldots, J$ for all t, s, 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 using the delta method from a block diagonal matrix formed from the covariance of the reduced-form VAR

coefficients, and the covariance obtained for H via MCMC.

B.5 GARCH parameters

In order to implement the hybrid GARCH 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 128 monthly macro variables from 1959 to 2018 included in McCracken & Ng's FRED-MD database. 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 128 series, standardized for comparison. 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(12) process. 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 7 displays the mean parameter values obtained across the series fitting into each of the sub-groups used by McCracken & Ng – first overall estimates, then each economic category. Note that for 22 of the series, the model could not successfully be fitted due to missing data. 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's simulation study, the overall estimates are the most relevant, as analysis proceeds on economy-wide factors. In addition, Figure 6 displays the distribution of estimated parameters for the overall dataset.

Table 7: *Estimates of GARCH(1,1) parameters for macroeconomic data*

Category	n	$\mu(1-\psi-\Upsilon)$	ψ	Υ
Overall	106	0.1097	0.6048	0.2476
Output and income	16	0.2870	0.3638	0.3207
Labor market	28	0.1251	0.5542	0.2683
Housing	4	0.0391	0.4243	0.1989
Consumption, orders, and inventories	4	0.3036	0.4034	0.1572
Money and credit	10	0.0989	0.5484	0.3470
Interest and exchange rates	21	0.0033	0.8028	0.1941
Prices	39	0.0253	0.7690	0.2017
Stock market	3	0.0600	0.8002	0.1435

Mean GARCH(1,1) parameters calibrated from AR(12) innovations from 128 monthly macro time series from McCracken & Ng FRED-MD database. The time series used are the residuals from an AR(12) estimated variable-by-variable on standardized data.

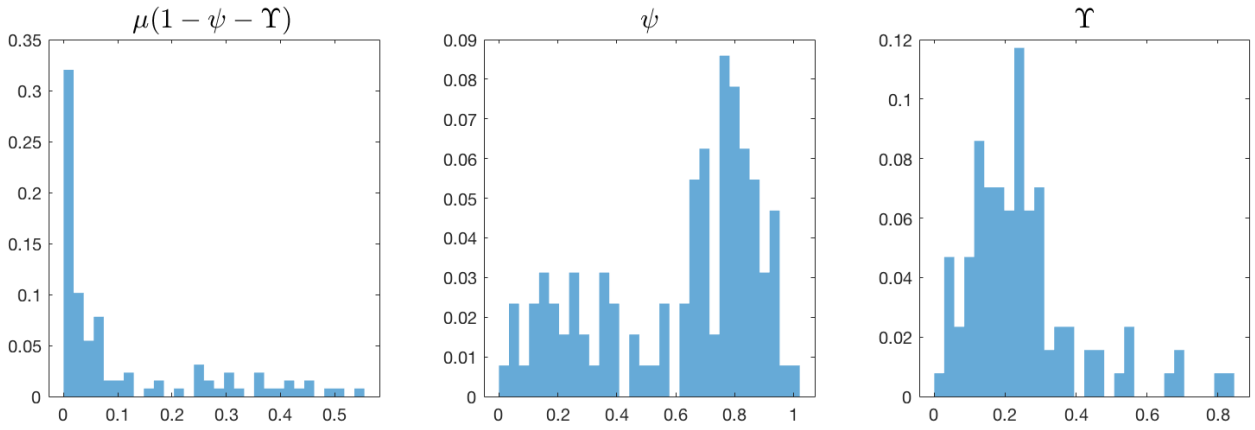

Figure 6: *Distribution of GARCH(1,1) parameters calibrated from AR(13) innovations 120 monthly macro time series. Categories follow those used in BBE. The time series used are the residuals from an AR(13) estimated variable-by-variable.*

C Online Appendix

Additional materials, including discussion of filtering algorithms, infill asymptotic approaches, and full simulation results, can be found at

https://scholar.harvard.edu/files/daniellewis/files/TVVID_OA.pdf