Identifying Shocks via Time-Varying Volatility

Daniel J. Lewis

Federal Reserve Bank of New York*

November 28, 2020

Abstract

I propose to identify an SVAR, up to shock ordering, using the autocovariance structure of the squared innovations, implied by an arbitrary stochastic process for the shock variances. These higher moments are available without parametric assumptions on the variance process. In contrast, previous approaches exploiting heteroskedasticity rely on the path of innovation covariances, which can only be recovered from the data under specific parametric assumptions on the variance process. The conditions for identification are testable. I compare the identification scheme to existing approaches in simulations and provide guidance for estimation and inference. I use the methodology to estimate fiscal multipliers peaking at 0.86 for tax cuts and 0.75 for government spending. I find that tax shocks explain more variation in output at longer horizons. The empirical implications of my estimates are more consistent with theory and the narrative record than those based on some leading approaches.

Keywords: identification, structural shocks, SVAR, fiscal multiplier, tax shocks, time-varying volatility, heteroskedasticity.

JEL codes: C32, C58, E20, E62.

*Email: daniel.lewis@ny.frb.org. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Helpful comments provided by Jim Stock, Gabriel Chodorow-Reich, Neil Shephard, Richard Crump, Mikkel Plagborg-Møller, Lutz Kilian, Jón Steinsson, Giri Parameswaran, Domenico Giannone, Davide Melcangi, Andrew Patton, participants at various seminars and conferences, as well as Dirk Krueger and four anonymous referees.

1 Introduction

There is compelling evidence of time-varying volatility in US macroeconomic data. The Great Moderation brought a marked decline in the volatility of output, inflation, and many other macro aggregates during the latter half of the twentieth century.[1] Financial time series also exhibit strong patterns of time-varying volatility, from business cycle to intraday frequencies.[2] The volatility of monetary policy shocks has also been found to vary over time, as has the volatility of tax revenues and government spending, so too economic policy uncertainty.[3]

I propose a new approach by which such time-varying volatility can be used to identify causal relationships in macroeconomic data. Econometricians typically seek to identify structural shocks, ε_t, from reduced form innovations to observable series, η_t, following the system of equations $\eta_t = H\varepsilon_t$. For example, a change in tax revenue could be driven by an exogenous tax shock, or an endogeneous automatic stabilizer effect of the business cycle. Up to second moments, there is no unique solution for H without further economic assumptions. I show how H may be identified from the autocovariance of squared reduced form innovations, η_t. This argument leverages statistical properties of the innovations implied by an arbitrary and unspecified stochastic process for the shock variances, rather than imposing economic restrictions. Identification holds without any parametric assumptions on the volatility process. In a simple model, this identification approach can be motivated as an instrumental variables problem; more generally, since structural shocks are assumed to be uncorrelated over time, the autocovariance of squared innovations picks up only dynamics of the volatility process, and I show that H can be recovered from these autocovariances given a rank condition on the structural variance process. I derive a test for this condition based only on the reduced form innovations. I use an application to fiscal policy to illustrate "identification via time-varying volatility" (TVV-ID), before investigating fiscal multipliers in detail.

The size of fiscal multipliers has been studied at length, under many identification approaches. However, the use of time-varying volatility to identify multipliers is relatively new (e.g., **?**, **?**). Considering a benchmark VAR from the literature, I document time variation in the squared residuals and find that the test for the identification condition is satisfied. This condition – on the rank of the structural variance process – at first appears solely statistical in nature, but has economic content. I argue that, provided that the individual structural shock variances are persistent, the rank condition will generally be satisfied if out-

[1]See e.g., **?** or **?**.

[2]See **?** and **?** for evidence at business cycle frequency and **?** for higher frequency data.

[3]**?** and **?** study monetary policy shocks, **?** fiscal policy, and **?** economic policy uncertainty.

put shock volatility does not predict tax shock volatility, consistent with classic treatments of tax shocks (e.g., **?**), and tax shock volatility does not predict output shock volatility, consistent with recent theoretical work by **?**. Moreover, these intuitive properties are satisfied by the shocks recovered using my identification approach and the classic **?** scheme, but at odds with those identified by **?**. I further show that the volatility profiles of shocks, which drive identification and are recovered as a byproduct, align with a narrative reading of US tax and output shocks. Finally, I illustrate that measuring second moments of the shocks is important since they help quantify the role of the shocks in explaining macroeconomic variation. While my identification is on face statistical, it draws on these properties that carry economic meaning.

I obtain peak values of 0.86 for tax multipliers and 0.75 for spending multipliers, with cumulative multipliers of 2.06 and 0.87 respectively after 20 quarters. While the response to tax shocks is delayed, they explain a greater share of output variation at longer horizons. My tax cut multipliers are lower than those of **?** and **?** because TVV-ID estimates a lower value of the tax automatic stabilizer effect, which is key to determining the size of the multiplier, as highlighted by **?**. The value I estimate, 1.58, is remarkably close to that calculated by **?** based on institutional data. I additionally characterize particular episodes that drive the identification of multipliers regardless of methodology, and show that the narrative record further favours my results based on TVV-ID, relative to existing approaches. By relating the statistical properties driving identification to familiar economic properties and events, and illustrating how they can help distinguish my estimates from existing findings, I provide a blueprint for how empirical researchers can unpack results based on my identification approach.

My identification argument contributes to a growing literature on statistical identification, both via heteroskedasticity and higher moments more generally. **?** and **?** share the insight that if the variances of the structural shocks change over time, shocks can be identified from the reduced-form covariances at different points in time. However, this path of reduced form covariances can be recovered by the econometrician only under specific parametric models. The method of **?** fits discrete variance regimes to the data, either based on external information or estimation. Generalizations have been made to Markov switching (**?**) and smooth transitions between regimes (**?**). **?** use the full path of covariances, recoverable from the data only under models like Generalized Autoregressive Conditional Heteroskedasticity (GARCH).[4] All of these approaches rely on these parametric features to consistently estimate

[4]While the identification argument is in principle non-parametric, based simply on a path of variances, this path can only be recovered from the data by an econometrician under functional forms like GARCH. These moments are thus not available to the econometrician, in the sense of being consistently estimable, without strong non-parametric assumptions. These distinctions are discussed in further detail in Section 2.4.

the path of variances, which has so far restricted researchers to choose one of the few models that can be accommodated. My identification approach is similar in spirit to identification based on non-Gaussianity (e.g., **???**), which uses different higher moments, but generally restricts any volatility processes of the shocks to be uncorrelated.

In contrast, my identification argument holds without parametric assumptions. While such a non-parametric argument naturally permits a non-parametric estimator, an equally important consequence is that it justifies the use of a wider range of parametric models, which may be more appealing in practice. Indeed, TVV-ID separately establishes identification via a novel channel for the models that have previously been shown to offer identification via heteroskedasticity. More importantly, it gives researchers the freedom to develop new alternative models and procedures to suit their data, without having to stop to establish identification from scratch. TVV-ID also admits models where the volatilities are state variables, as in the stochastic volatility (SV) model. The parameters of interest can still be consistently estimated in such models, since identification follows from moments that can be consistently estimated even when the volatilities cannot. Thus, while a non-parametric generalized method of moments (GMM) approach is a natural implementation of TVV-ID, the researcher is also empowered to use a (quasi-) likelihood approach based on any model that implies an autocovariance for the squared innovations.

I provide guidance on the use of TVV-ID in practice. The test for the identification conditions performs well in simulations. I also compare estimators applying existing statistical identification approaches to newly-admissible estimators based on TVV-ID in simulations. I find that one such new estimator, based on a first-order autoregressive (AR(1)) SV model, performs best across many data-generating processes (DGPs), and I apply it in my empirical analysis. I provide a formal framework for statistical inference on H, accounting for the fact that identification holds only up to column order. I put these results together to provide step-by-step guidance for empirical researchers, spanning testing the identification conditions, estimating H, conducting inference on H, and extending inference to impulse response functions (IRFs).

The remainder of this paper proceeds as follows. Section 2 describes the identification problem in detail and presents the theoretical results. Section 3 describes estimation and inference procedures, providing a step-by-step guide for empirical use. I study fiscal multipliers in Section 4. Section 5 concludes.

Notation

The following notation is used in the paper. $\otimes$ represents the Kronecker product of two matrices; $\odot$ represents the element-wise (Hadamard) product of two matrices; $B_{(i)}$ denotes the i^{th} row of matrix B; $B^{(j)}$ denotes the j^{th} column of matrix B; B_{ij} denotes the ij^{th} element of matrix B; $B^{(-i)}$ denotes all columns of B except for the i^{th}, and similarly for rows and elements; $diag\,(b)$ is a diagonal matrix with the vector b on the diagonal; $id\,(B)$ inverts the matrix diagonal of the square matrix B; $x_{1:t}$ denotes $\{x_1, x_2, \ldots, x_t\}$. A permutation matrix P_i is one of the set of $n!$ matrices consisting of a reordering of the columns of I_n.

Additionally, I use the non-standard notation $E_t\,[\cdot]$ to explicitly denote a time-specific expectation, i.e. the mean value of x_t at time t, as opposed to across t, and similarly $E_{t,s}\,[\cdot]$ when both time t, s variables are contained in the argument. This notation is used to clarify when stationarity is not assumed, 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.

2 Identification theory

In the canonical SVAR setting, a vector of innovations, η_t, is composed of unobserved orthogonal structural shocks, ε_t, via a response matrix, H. η_t is an $n \times 1$ vector, obtained from a VAR of observables Y_t as $A\,(L)\,Y_t = \eta_t$, where $A\,(L)$ is a lag polynomial. Similarly, ε_t is an $n \times 1$ vector, and H an $n \times n$ matrix. Thus,

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

leaving H and, equivalently, ε_t, to be identified.[5] I begin by presenting a simple example under special assumptions to outline the identification problem and how heteroskedasticity may solve it. I then derive a representation of higher moments of the reduced-form innovations to serve as identifying equations in a fully general model. I go on to establish conditions under which these equations have a unique solution. I highlight the role of the various assumptions and identification conditions, propose a simple test of those conditions, and explain the relation to existing identification approaches.

2.1 Intuition for the use of heteroskedasticity

Before the impact of heteroskedasticity can be illustrated, some standard assumptions underlying equation (1) are required.

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

[5] Equation (1) can alternatively describe any other decomposition into orthogonal components, where second moments offer identification only up to orthogonal rotations, including factor models, for example.

1. $E\left[\varepsilon_t \varepsilon_t' \mid \sigma_t\right] = diag\left(\sigma_t^2\right) \equiv \Sigma_t$ *(σ_t^2 is the conditional variance of the shocks),*

2. σ_t *is a strictly positive stochastic process and* ε_t *has finite fourth moments,*

3. $E\left[\Sigma_t\right] = \Sigma_\varepsilon$,

4. *Shocks satisfy conditional mean independence,* $E\left[\varepsilon_{it} \mid \varepsilon_{is}\right] = 0, s \neq t$ *and* $E\left[\varepsilon_{it} \mid \varepsilon_{-is}\right] = 0$ *for all* i*, all* $t, s = 1, 2, \ldots T$,

5. H *is time-invariant, full rank, and has a unit diagonal.*

The fourth point substitutes conditional mean independence for the usual slightly weaker uncorrelated shocks assumption. While the variance of shocks may change, fixing H (as in Assumption 0.5) means that the economic impact of a unit shock remains the same. It is natural to seek to identify H from the overall covariance of η_t, $E\left[\eta_t \eta_t'\right] = \Sigma_\eta$. However, it is well-known that these equations can only identify H up to an orthogonal rotation, Φ (where $\Phi\Phi' = I$).[6]

Variation in Σ_t may allow the researcher to overcome this indeterminacy. Consider a simple two-variable example, where one structural variance is time-varying and the other is fixed. For example, in the fiscal setting, the model could simply contain tax revenues and output, with innovations η_{Tt}, η_{Yt} and shocks $\varepsilon_{Tt}, \varepsilon_{Yt}$. Assume that the tax shock's volatility is fixed (if tax shocks capture ideologically-motivated non-cyclical changes in tax policy, those dynamics may be relatively stable from the 1950s-2000s), while the output shock's volatility varies (capturing the Great Moderation, for example). In this simplified setup, the **?** approach yields closed form solutions for H (as in **?**, for example). Ordering tax first and output second, with $\sigma_{T,t}^2 \equiv \sigma_T^2$, denote

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

H_{12} will be the parameter of interest, representing the automatic stabilizer effect (this is the key parameter highlighted by **?** and calibrated by **?**). 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, for example, the period before the Great

[6] Observe $\Sigma_\eta = H\Sigma_\varepsilon H' = (H\Phi)(\Phi'\Sigma_\varepsilon\Phi)(H\Phi)' = H^*\Sigma_\varepsilon^* H^{*'}$, where $H^* = H\Phi D_{H,\Phi}$ and $\Sigma_\varepsilon^* = D_{H,\Phi}^{-1}\Phi'\Sigma_\varepsilon\Phi D_{H,\Phi}^{-1}$, with $D_{H,\Phi} = id\,(H\Phi)$ 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, or H is identified up to transformations $H\Phi \times id\,(H\Phi)$. Alternatively, note that due to the symmetry of Σ_η, it offers $n\,(n+1)/2$ equations, but there are n^2 unknowns. This is the fundamental identification problem posed by the SVAR methodology and indeed many related models (e.g., factor models).

Moderation, and the Great Moderation, it is shown by **?** (and in the Supplement) that

$$\frac{E_{T_A}\left[\eta_{Tt}\eta_{Yt}\right]-E_{T_B}\left[\eta_{Yt}\eta_{Yt}\right]}{E_{T_A}\left[\eta_{Yt}^2\right]-E_{T_B}\left[\eta_{Yt}^2\right]}=\frac{H_{12}\left(E_{T_A}\left[\sigma_{Y,t}^2\right]-E_{T_B}\left[\sigma_{Y,t}^2\right]\right)}{E_{T_A}\left[\sigma_{Y,t}^2\right]-E_{T_B}\left[\sigma_{Y,t}^2\right]}=H_{12}, \tag{2}$$

Assuming that $\Delta\left(\sigma_{Y,t}^2\right)\equiv E_{T_A}\left[\sigma_{Y,t}^2\right]-E_{T_B}\left[\sigma_{Y,t}^2\right]\neq 0$ (the output shock variance *did* change during the Great Moderation), H_{12} can thus be identified in closed form. ε_t need only have finite second moments for all $t\in T_A, T_B$. While this argument is motivated by a regime-based process, identification holds even when misspecified, provided $\Delta\left(\sigma_{Y,t}^2\right)\neq 0$ and σ_T^2 is indeed fixed. If there are regimes, they need not be known or correctly specified, as noted in **?**. However, if the value of the $\sigma_{Y,t}^2$ process is instead constant, $\Delta\left(\sigma_{Y,t}^2\right)$ is zero in population, and identification fails.

The preceding approach exploits moment conditions based on subsample means of the variance process. I now break with **?** and offer an entirely novel argument, based instead on the full-sample autocovariance properties of $\eta_t\eta_t'$, illustrating the intuition of TVV-ID. In fact, exploiting the autocovariance can be closely related to an instrumental variables approach. Noting

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

it is clear that H_{12} would be identified from the ratio of the $H_{12}\varepsilon_{Yt}^2$ and ε_{Yt}^2 terms. This is not feasible as only the values of η_t are observed, and not their separate components. However, a lagged value of η_{Yt}^2 can be used as an instrument for ε_{Yt}^2. Note

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

by Assumption 0.4 and the fact that the variance of tax shocks, σ_T^2, is fixed, so $cov\left(\varepsilon_{Tt}^2,\varepsilon_{T(t-p)}^2\right)=0$. Now, instead of relying on the fact that the variance of the output shock changes across subsamples, I exploit the fact that it is *persistent* (while varying). It may still decline on average between 1950 and 2000, but I no longer need to specify such external information. H_{12} is then identified in closed form:

$$\frac{cov\left(\eta_{Yt}\eta_{Tt},\eta_{Y(t-p)}^2\right)}{cov\left(\eta_{Yt}^2,\eta_{Y(t-p)}^2\right)}=\frac{H_{12}cov\left(\varepsilon_{Yt}^2,\varepsilon_{Y(t-p)}^2\right)}{cov\left(\varepsilon_{Yt}^2,\varepsilon_{Y(t-p)}^2\right)}=H_{12}. \tag{3}$$

This is the familiar instrumental variables estimator, where the dependent variable is $\eta_{Yt}\eta_{Tt}$, the endogenous regressor is η^2_{Yt}, and the instrument is $\eta^2_{Y(t-p)}$. The idea of instrumenting for ε^2_{Yt} is related to the identification argument of **?** for cross-sectional settings, who exploits the existence of some instrument for the heteroskedasticity; equation (3) also takes a similar form to the panel identification results of **?**. In the present uniquely time series argument, equation (3) produces a closed form solution for H_{12} because the previous value $\eta^2_{Y(t-p)}$ is uncorrelated with all period t terms except those containing ε^2_{Yt}. H_{12}, the automatic stabilizer effect, is identified as a coefficient on the persistence of the output shock variance, without supplying an external calibration (**?**), an external instrument (**?**), or exogenous regime dates (the **?** approach). The value of H_{21}, the instantaneous impact of tax shocks on output, is then easily identified from $E\left[\eta_t\eta_t'\right] = H\Sigma_t H'$. The identification argument applies for any lag, p. Identification holds provided

$$cov\left(\varepsilon^2_{Yt}, \varepsilon^2_{Y(t-p)}\right) \neq 0$$

for some p.

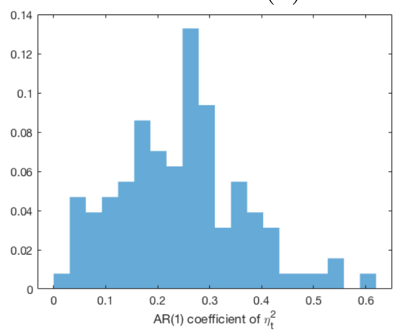
This requirement that the p^{th} autocovariance of ε^2_{Yt} is non-zero is satisfied by a variety of processes for σ^2_{Yt}. If the true process is regime-based, as suggested by the Rigobon estimator, identification follows from the non-zero autocovariance around break dates. In an SV model, it holds if the autoregressive coefficient is non-zero. In a GARCH model, at least one of the autoregressive parameters must be non-zero. For the fiscal multipliers model, I find in Section 4 that the variance of the output shock is indeed highly persistent. This simple example illustrates 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 specification – is not shared by the existing approaches. I have made no assumptions about whether the heteroskedasticity is conditional or unconditional (either can imply a suitable autocovariance) and I have required only that the volatility process exhibits some degree of persistence.

Empirically, there is strong evidence of such persistence in macro time series, as discussed in **?**, for example. As a simple exercise, Figure 1 displays AR(1) parameters of η^2_t, where η_t are residuals of univariate AR(12) models fitted to each series of the **?** FRED-MD database in turn. I reject the null hypothesis that the AR(1) coefficient is zero at the 1% level for 96 of the 128 series, 5% for 98, and 10% for 101. A Ljung-Box test, as adopted by **?**, rejects homoskedasticity at the 1% level for 100 of the series and the 5% level for 103. The persistence required of the squared shocks for identification is present in many of these squared innovations.

In this simple case, multiple autocovariances can easily be combined; each yields moments

Figure 1: Distribution of 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 the **?** FRED-MD's 128 monthly time series. The figure displays the distribution of the implied AR(1) coefficients of η_t^2.

of the form

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

which can be stacked to yield an overidentified GMM problem. Alternatively, it might be natural to assume that the (log) variances follow some loose parametric form, like an AR(1), and let this assumption imply a whole range of autocovariances.

2.2 Identification via time-varying volatility

In the previous section, I made strong assumptions to assist intuition. I now relax them and develop TVV-ID in its general form. I consider an $n-$dimensional model and allow the variance of all shocks to vary over time. Through the lens of the fiscal example, government spending can now be included with tax revenue and output, and the variance of all three shocks, not just the output shock, can vary over time. While the identification argument becomes more complicated, not only are these richer dynamics necessary to realistically model empirical settings, but they offer valuable identifying information; in Section 4.3 I find that the intertemporal relationship of these variances *across* shocks is instrumental in distinguishing the fiscal multipliers obtained by TVV-ID from those of **?**.

Again, let

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

Denote $\mathcal{F}_{t-1} = \{\varepsilon_1, \dots \varepsilon_{t-1}, \sigma_1, \dots \sigma_{t-1}\}$. I replace Assumption 0 with Assumption A:

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

1. $E_t\left(\varepsilon_t \mid \sigma_t, \mathcal{F}_{t-1}\right) = 0$ *and* $E_t\left(\varepsilon_t\varepsilon_t' \mid \sigma_t, \mathcal{F}_{t-1}\right) = \Sigma_t$,

2. $\Sigma_t = diag\left(\sigma_t^2\right), \sigma_t^2 = \sigma_t \odot \sigma_t$,

3. $E_t\left[\sigma_t^2\right] < \infty$.

These assumptions cover both SV and autoregressive conditional heteroskedasticity-type (ARCH) models (where σ_t is a function of $\varepsilon_1, \ldots \varepsilon_{t-1}$), amongst many others, including unconditional heteroskedasticity. By explicitly conditioning on σ_t, this setup rules out innovations to the σ_t process being correlated with ε_t, as is true of VAR models more broadly.

In addition, I make standard assumptions on H:

Assumption B. *H is time-invariant, full rank, and has a unit diagonal.*[7]

The assumption that H is time-invariant is crucial for identification and ubiquitous in the literature. In fact, identification has not been established when H is time-varying except in very special cases (e.g., **???**). Work allowing more flexible time-variation in H is limited to Bayesian frameworks, most notably **?** and **?**. I discuss the time-invariance of H further following the main identification results. Implicit in this setting and most related work is the additional assumption of invertibility, so the shocks are fundamental and thus recoverable from η_t.

Decomposition of $\eta_t \eta_t'$

To obtain moments in terms of just H and the underlying volatility process, I work with a transformation of η_t, (ζ_t, defined below), as my basic data. I begin by writing the decomposition

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

where Σ_t is unknown. Define L to be an elimination matrix, and G a selection matrix (of ones and zeros), see **?**, for example.[8] Then

$$\begin{aligned}
\zeta_t &\equiv vech\left(\eta_t \eta_t'\right) = vech\left(H \Sigma_t H'\right) + vech\left(V_t\right) \\
&= L\left(H \otimes H\right) vec\left(\Sigma_t\right) + v_t, \;\; v_t = vech\left(V_t\right) && (4) \\
&= L\left(H \otimes H\right) G \sigma_t^2 + v_t, && (5)
\end{aligned}$$

[7]The unit diagonal assumption is simply a normalization. Note that even if there are zeros in H, such that certain column orderings are incompatible with a unit-diagonal, this poses no problem for *identification* of the columns of H, since column order is imposed only ex-post for interpretation. The conventional alternative, a unit–variance normalization, makes less sense in this setting. First, instead of n unit variances, it would impose a unit normalization on some n elements of the $n \times n^2$ matrix $M_{t,s}$, defined below, a nuisance matrix encoding a variety of information about the persistence properties of the variances. Given that this is not a familiar object, nor necessarily easily interpretable, it is not a natural target for normalization. Second, as I discuss in Section 3 (and apply empirically), it may be desirable to estimate H using several alternative models (or no model) for the variances. These results will only be directly comparable if the normalization is consistent, but H is the only object that in general will be common across such estimators.

[8]This means $vech\left(A\right) = Lvec\left(A\right)$ and $vec\left(ADA'\right) = \left(A \otimes A\right) Gd$ where $d = diag\left(D\right)$.

The simplification from (4) to (5) in the first term follows due to the diagonality of Σ_t under Assumption A.2. From the definition of V_t and Assumptions A.1, A.3, and B, $E_t\left[V_t \mid \sigma_t, \mathcal{F}_{t-1}\right] = 0$, 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 A.3 that I can integrate over Σ_t to obtain $E_t\left[v_t \mid \mathcal{F}_{t-1}\right] = 0$ and similarly that $E_t\left[|v_t|\right] < \infty$. Therefore v_t is a martingale difference sequence. Each observation of ζ_t is an observation of $H\Sigma_t H'$, plus mean-zero noise.

Properties of ζ_t

Assumption C expands on Assumption A.3 to permit a characterization of the autocovariance of ζ_t.

Assumption C. For every t, $E_t\left[vec\left(\varepsilon_t\varepsilon_t'\right) vec\left(\varepsilon_t\varepsilon_t'\right)'\right] < \infty$.

Using this additional assumption, the autocovariance of ζ_t has a convenient form:

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

$$cov_{t,s}\left(\zeta_t, \zeta_s'\right) = L\left(H \otimes H\right) G M_{t,s}\left(H \otimes H\right)' L', \; t > s \tag{6}$$

where

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

This equation represents an reduced form quantity, $cov_{t,s}\left(\zeta_t, \zeta_s'\right)$, as a product of H and the $n \times n^2$ matrix $M_{t,s}$ (composed of $n \times \left(n^2+n\right)/2$ moments of the underlying variance process). If $E_{t,s}\left[\sigma_{it}^2\left(\varepsilon_s\varepsilon_s' - \Sigma_s\right)\right]$ is diagonal (as in an SV model, or if any ARCH effects come from only squared shocks), $M_{t,s}$ can be replaced with $M_{t,s}^{red}G'$ where $M_{t,s}^{red}$ is only $n \times n$.

An autocovariance of the vectorization of $\eta_t\eta_t'$ can thus be expressed as just a product of H, an $n \times n^2$ nuisance matrix, and known matrices of zeros and ones. This is remarkably parsimonious for a covariance of random matrices. Note that stationarity has not been assumed, merely the existence of higher moments. 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. A single autocovariance provides $(n^2+n)/2 \times (n^2+n)/2$ equations in $n^2 - n + n(n^2+n)/2$ parameters, so the order condition is satisfied.

A unique solution

Having derived a set of equations of adequate order to identify H, it remains to show that they yield a unique solution. Conditions under which (6) yields a unique solution for H are established by Theorem 1.

Theorem 1. *Under Assumptions A.1-2, B, & C, equation (6) holds. Then H and $M_{t,s}$ are jointly uniquely determined from (6) (up to column order) provided* $\text{rank}(M_{t,s}) \geq 2$ *and $M_{t,s}$ has no proportional rows.*

Theorem 1 states that (under certain conditions) equation (6) will yield a unique solution for the relative magnitudes of elements in each column of H, with normalization imposed by Assumption B. The identification result is based on period-specific moments – an autocovariance between two specific time periods, s, t – so stationarity is neither assumed nor required. For example, a stochastic volatility process with a unit root could in principle be used to identify shocks by TVV-ID. In practice though, fourth-order stationarity of ε_t will often be assumed so that (6) may be consistently estimated across the full sample (although this need not be the case in an infill-asymptotic framework). The solution is unique up to column order, or equivalently choosing economic labels for the shocks that dictate an ordering given the unit diagonal normalization imposed on H by Assumption B (for example, labeling a shock as the tax shock, with tax revenues being the first variable in the VAR, dictates the corresponding column of H should be ordered first).[9] There are $n!$ such orderings; the same is true for any statistical identification approach, including those based on heteroskedasticity or non-Gaussianity. Labeling the columns and selecting an associated order, to complete global identification, is needed to render the shocks interpretable in an economic sense, and also important for statistical inference, as discussed in Section 3.2.

Theorem 1 makes two requirements of $M_{t,s}$. First, it must have rank of at least 2. Second, it must have no proportional rows. This is weaker than a full rank condition, since rows of $M_{t,s}$ may be linear combinations, as long as they are not simply proportional. This dimensionality requirement ensures adequate heterogeneity in $M_{t,s}$ to uniquely identify H. $M_{t,s}$ encodes the autocovariance (and potentially ARCH) properties of σ_t^2 with σ_s^2; in an SV model, $M_{t,s}$ is simply the autocovariance of σ_t^2 multiplied by G'. These conditions imply that all n variances must be time-varying and persistent, and additionally that no two variance

[9] In other words, H and $HP_i \times id(HP_i)$, where P_i is a permutation matrix, are observationally equivalent; see Section 3.2 for a detailed discussion.

processes can have fully proportional autocovariance structures with respect to σ_s^2. This will not occur if each variance has at least some persistent idiosyncratic component, except in knife-edge cases. **?** find that there are indeed strong idiosyncratic components in time-varying volatility that cannot be explained by common factors; the identification conditions will hold if those idiosyncratic components are persistent, and thus impact $M_{t,s}$. For fiscal multipliers, as I discuss in Section 4.3, theoretical arguments based on the **?** notion of a tax shock and the results of **?** suggest that the co-persistences of tax and output shock variances should be zero. Since both variances are strongly persistent, this implies that the tax and output rows of $M_{t,s}$ are linearly independent, satisfying the rank condition. Unless the spending shock variance is almost perfectly predicted by one of $\sigma^2_{T,t-1}$, $\sigma^2_{Y,t-1}$ but not at all by the other, the $\sigma^2_{G,t}$ row also cannot be proportional to either of those rows, so identification holds.

If the proportional row condition on $M_{t,s}$ does in fact fail, partial identification is still possible, as established in Corollary 1.

Corollary 1. *Under Assumptions A.1-2, B, & C, equation (6) holds. Then $H^{(j)}$ is identified from (6) provided $\text{rank}\left(M_{t,s}\right) \geq 2$ and $M_{t,s}$ contains no rows proportional to row j.*

This shows that columns of H pertaining to shocks whose volatility processes do not have proportional autocovariance structures can still be identified.

The identification conditions in Theorem 1 can be relaxed by exploiting additional identifying equations. If, for example, the usual covariance of residuals,

$$E_t\left[\eta_t \eta_t'\right] = E_t\left[\zeta_t\right], \tag{7}$$

is considered, Theorem 1 can be supplanted by Theorem 2.

Theorem 2. *Under Assumptions A.1-2, B, & C, equation (6) holds. Then H is uniquely determined from (6) and (7) (up to column order) provided $\left[\begin{array}{cc} M_{t,s} & E_t\left[\sigma_t^2\right] \end{array}\right]$ has rank of at least 2 and no proportional rows.*

Theorem 2 shows that, if the covariance of η_t is also used as an identifying moment, a proportional row assumption must additionally relate $E_t\left[\sigma_t^2\right]$ to $M_{t,s}$ in order for identification to fail. Similar arguments can be made, adding further observable moments, requiring proportionality extend to a matrix with progressively more columns. A major implication of Theorem 2 is described in Corollary 2.

Corollary 2. *H is uniquely determined from (6) and (7) (up to column order) if at least $n-1$ shocks display time-varying volatility with non-zero autocovariance, provided that for no two shocks i, j, $cov_{t,s}\left(\sigma_{it}^2, vec\left(\varepsilon_s \varepsilon_s'\right)'\right) = cov_{t,s}\left(\sigma_{jt}^2, vec\left(\varepsilon_s \varepsilon_s'\right)'\right) \frac{E_t\left[\sigma_{it}^2\right]}{E_t\left[\sigma_{jt}^2\right]}$.*

Corollary 2 states that with the addition of (7), only $n-1$ dimensions of persistent time-varying volatility are sufficient to identify H, except in a very special case. This degenerate case amounts to the autocovariance structure of two shock variances (with respect to *every* $\varepsilon_{is}\varepsilon_{js}$) being proportional, with the proportionality constant just happening to be the ratio of their means. As discussed in Section 2.4, this weaker identification condition puts TVV-ID on a level theoretical footing with existing heteroskedasticity-based approaches. In Section 2.3, I propose a formal test to evaluate Theorem 2's identification conditions in practice.

All of the preceding identification results yield identification of H up to column order, and it remains to label the shocks in a manner that dictates some ordering, or otherwise choose an ordering. Once an order is determined, normalization (scaling) follows from Assumption B. I discuss in Section 3.2 how with the use of additional economic information, point identification holds globally, and consider the implications for inference on estimates of H.

Time-invariance of H

While TVV-ID focuses on the instability of the variances of structural shocks, H is assumed fixed. Although this is in principle a strong assumption, no existing identification scheme can flexibly accommodate time-varying H (although **?**, **?**, **?**, and **?** do so under very specific functional forms). Even the simplest recursive short-run restrictions do not identify a known moment of H if H is in fact time-varying. Allowing H to vary more generally presents an interesting econometric problem, which warrants further study. While there are workhorse models in macroeconomics that allow for time-varying H, (e.g., **??**), these all adopt a Bayesian framework without identification results to separate variation in H from variation in Σ_t based on properties of the observable data alone. In this context, the parameter values obtained are driven by the structure of the priors, imposing information the data could never offer. As such, these approaches are largely orthogonal to the goal of this paper to provide non-parametric frequentist identification results facilitating consistent estimation of H based on observable data and (relatively) mild assumptions. While some frequentist work has adopted time-varying parameters (TVP) in the reduced form model, for example **?**, such papers are still unable to incorporate variation in H; time-variation in reduced form parameters can be combined with TVV-ID.

Still, there are two ways in which time-variation in H is potentially compatible with TVV-ID. First, if H varies at a slower rate than the variances, identification may still hold asymptotically; H will be locally stationary over intervals over which the variances are not. Such a case could be explored in an infill-asymptotic setting, for example. Theoretical work sometimes reflects such distinctions in the rate of variation; for example, **?** split volatility into short-run and long-run components, with agents' behaviour driven by the slower moving component. Second, compared to identification exploiting regimes, as in **?**, TVV-ID is better equipped to permit estimation using sub-samples over which H may plausibly be fixed, since the data do not need to be subdivided again for identification under a constant H. Should a researcher remain worried about the assumption of a fixed H, tests of overidentifying restrictions remain an option, as H may be over-identified by TVV-ID. Further, **?** 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.[10]

2.3 Testing the identification conditions

Testing conditions for identification based on heteroskedasticity is difficult in general. The requirements for identification impose conditions on parameters that are only identified *conditional* on identification holding. In **?**, the time paths of structural variances are required to be linearly independent, and in **?** the two (or more) sets of structural variances must be non-proportional. In TVV-ID, $M_{t,s}$ or $\left[\begin{array}{cc} M_{t,s} & E_t\left[\sigma_t^2\right] \end{array} \right]$ must have rank of at least 2 and no proportional rows. Given knowledge of the structural parameters, these conditions could easily be tested, but those parameters cannot be recovered without assuming identification. As a result, **?** propose a test for the dimension of heteroskedasticity based on the autocovariance of $vech\left(\varepsilon_t\varepsilon_t'\right)$, but must first impose a recursive structure to recover ε_t (so as not to assume H is identified via heteroskedasticity), and so base a test on some orthogonal rotation of the true structural shocks. In contrast, I derive implications of the identification conditions in Theorem 2 that are testable based on reduced form moments alone. In this section, I assume stationarity to simplify asymptotic results (and so replace $\zeta_t\zeta_s'$ with $\zeta_t\zeta_{t-p}'$, etc.).

Define $\tilde{M}_{t,t-p} = E\left[\sigma_t^2 vec\left(\varepsilon_{t-p}\varepsilon_{t-p}'\right)'\right]$. Proposition 2 provides testable implications of the rank of $\tilde{M}_{t,t-p}$ for the reduced form moment $E\left[\zeta_t\zeta_{t-p}'\right]$ and relates the rank of $\tilde{M}_{t,t-p}$ to the identification conditions. **?** propose a similar rank test for variance components applied to reduced-form properties (a lag$-l$ generalized kurtosis matrix, like $cov\left(\zeta_t, \zeta_{t-p}'\right)$), but their results assume variances follow an ARCH functional form.

Proposition 2. *By construction,* $rank\left(E\left[\zeta_t\zeta_{t-p}'\right]\right) = rank\left(\tilde{M}_{t,t-p}\right) = r$*; if* $r = n$*,* $\tilde{M}_{t,t-p}$ *is full rank and the identification conditions of Theorem 2 are satisfied.*

[10]The less-familiar assumptions needed in **?**, those of Near-Epoch Dependence (NED), can be replaced by stronger properties that hold for both GARCH and SV processes. **?** shows that GARCH satisfies β-mixing (and thus α-mixing with exponential rate) and **?** show that SV models inherit the mixing properties of the log-variance process. Results in **?** show that an AR(1) variance process is $\alpha-$mixing with exponential rate. These mixing properties can be shown to imply NED; see **?** Chapter 17 for additional background.

A similar proposition holds for $cov\left(\zeta_t, \zeta_{t-p}'\right)$ and $M_{t,t-p}$ to verify the conditions of Theorem 1. However, estimating the centered moment $cov\left(\zeta_t, \zeta_{t-p}'\right)$ combines estimates of both the uncentered higher moments, $E\left[\zeta_t\zeta_{t-p}'\right]$, and $E\left[\zeta_t\right]$, introducing additional estimation error relative to simply working with $E\left[\zeta_t\zeta_{t-p}'\right]$. In small samples, this decrease in precision can negatively impact tests based on the estimated moments.[11] Additionally, a test based on the rank of $\tilde{M}_{t.t-p}$ will have power against the null of non-identification in several leading cases where one based on the rank of $M_{t,t-p}$ will not. In particular, this is true when one shock exhibits homoskedasticity (in which case the corresponding row of $M_{t,t-p}$ is zero) or when one variance is a linear transformation (with a non-zero intercept) of another (in which case the corresponding rows of $M_{t,t-p}$ are proportional); in contrast, $\tilde{M}_{t,t-p}$ remains full rank due to the different mean variances.

The implication of $r = n$ for $\tilde{M}_{t,t-p}$ is actually stronger than the condition required for identification in Theorem 2 in two ways. As detailed in the remark preceding the proof of Proposition 2, $rank\left(\tilde{M}_{t,t-p}\right) \leq rank\left(\left[\begin{array}{cc} M_{t,t-p} & E\left[\sigma_t^2\right]\end{array}\right]\right)$, although equality holds except in certain nonlinear models. Second, the identification results require only a rank of 2, with no proportional rows (rows that are not proportional but are otherwise linear combinations lower the rank of $\tilde{M}_{t,t-p}$ and $\left[\begin{array}{cc} M_{t,t-p} & E\left[\sigma_t^2\right]\end{array}\right]$ but do not prevent identification). Thus, the condition $rank\left(E\left[\zeta_t\zeta_{t-p}'\right]\right) = n$ can be viewed as conservative with respect to the true identification conditions for TVV-ID. The extent to which this matters in practice is an empirical question – it is not a barrier in this paper's empirical study – and one left to future work.[12]

The problem of testing for identification is now reduced to testing the rank of $E\left[\zeta_t\zeta_{t-p}'\right]$. Tests of matrix rank have been studied extensively, for example by **?**. I assume the availability of a consistent and asymptotically normal estimator for $E\left[\zeta_t\zeta_{t-p}'\right]$ (e.g., $\frac{1}{T-1}\sum\hat{\zeta}_t\hat{\zeta}_{t-p}'$); one set of conditions assuring this is that ε_t is M-dependent and is eighth-order stationary, for example. Then, Theorem 3 provides a test statistic and asymptotic distribution to assess the rank of $E\left[\zeta_t\zeta_{t-p}'\right]$, and thus test whether the conditions to identify H using TVV-ID hold.

Theorem 3. *If $E\widehat{\left[\zeta_t\zeta_{t-p}'\right]}$ is an asymptotically normal estimator of $E\left[\zeta_t\zeta_{t-p}'\right]$, then under the null hypothesis that the matrix has rank r, the associated Cragg-Donald statistic $CD_{\zeta,p}(r)$ has the asymptotic distribution $CD_{\zeta,p}(r) \xrightarrow{d} \chi^2\left(\left(\left(n^2+n\right)/2 - r\right)^2\right)$.*

The interested reader should consult **?** for additional technical details and a description

[11]Indeed, I find in unreported simulations that performance is better for tests of the rank of $\tilde{M}_{t,t-p}$ than $M_{t,t-p}$.

[12]As detailed at the end of Section 2.4, this also does not appear to be a problem in other empirical settings.

of the test statistic. Essentially, the test assesses the deviation of the lower right part of the estimated matrix from zero following r steps of Gaussian elimination.

I explore the performance of the test for identification described by Proposition 2 and Theorem 3 in a simulation study reported in Section 4 of the Supplement. I find that the test exhibits good size control even in small samples. In the vast majority of specifications, the tests are very well-sized, or even conservative; for only one DGP are size distortions over 2% (for a 5% nominal test). The test also offers reasonable (size-adjusted) power. **?** conduct simulations based on the **?** test, assessing the autocovariance of $vech\,(\varepsilon_t\varepsilon_t')$, instead of $vech\,(\eta_t\eta_t')$, which is the object of my test. However, as noted above, $vech\,(\varepsilon_t\varepsilon_t')$ cannot be recovered without imposing a value for H, which they do by assuming a Cholesky structure (since heteroskedasticity is being tested, so cannot be assumed). Thus, one structural model is assumed to test for variation that would allow the researcher to ultimately identify a different structural model. This additional assumption makes direct comparison of the tests difficult, but their test does appear to offer superior power properties. However, further simulations show that my test also performs considerably better in this dimension when applied to (empirically infeasibly) orthogonalized structural shocks instead of reduced form innovations.

2.4 Relation to existing approaches

TVV-ID generalizes the conditions under which previous approaches have established identification via heteroskedasticity and nests the parametric models on which they have relied. Below, I describe the relation of TVV-ID to each of the existing identification results.

? offer an identification argument that is in principle non-parametric; they show that, conditional on the time path of reduced form covariances, $\Sigma_{\eta 1:T}$, H is identified, provided the variance processes are linearly independent. However, this path is not in general available to the econometrician, who observes only the noisy $\eta_t\eta_t'$ in each time period, no matter the sample length. This leads the authors to recommend a GARCH functional form, which is very special in allowing the reduced form covariances to be deterministically recovered from the observations conditional on H and the parameters of the variance process. This property is shared by virtually no other variance process.

TVV-ID avoids these issues entirely by not making any reference to the variance path for identification, instead using unconditional moments, in particular the autocovariance of ζ_t. Because it is unnecessary to recover the variance path for identification, TVV-ID can admit a near arbitrary range of volatility models, and is truly non-parametric. Such moments can, under suitable assumptions, be consistently estimated even in models with state variables.

TVV-ID is the first scheme to imply that H can be consistently estimated *even when the volatility path cannot.* **?** exploit this result to estimate a stochastic volatility model for the structural shock variances, very closely related to the model I discuss in Section 3 and employed in Section 4. Moreover, TVV-ID explicitly nests the GARCH implementations of **?** because a (stationary) GARCH process clearly implies a suitable matrix $M_{t,t-p}$ for autocovariance p.[13]

? regime-based identification following from subsample, Markov switching (e.g., **?**), or smooth transition (e.g., **?**) models are also nested by TVV-ID. The transitions between regimes imply non-zero autocovariance for the shock variances.[14] Both **?** and **?** arguments require $n-1$ dimensions of linearly independent time-varying volatility; TVV-ID similarly requires $n-1$ volatility processes with non-proportional autocovariance structures (Corollary 2). The TVV-ID condition has the advantage that it may be satisfied even if the variances are, on average, proportional across regimes, since there may be differences in autocovariance properties from one observation to the next that TVV-ID can exploit.

TVV-ID belongs more broadly to a long literature (dating to at least **?** and **?**) of identification based on higher moments. While TVV-ID exploits intertemporal moments, work has generally focused on contemporaneous moments (or cumulants), either via maximum likelihood (e.g., **??**) or independent components analysis (ICA) (e.g., **?**). If $n-1$ shocks are non-Gaussian, identification holds, again parelling the dimensionality requirement of Corollary 2. Since heteroskedasticity induces unconditional non-Gaussianity even if shocks are conditionally Gaussian, in theory identification via heteroskedasticity is nested by arguments based on non-Gaussianity. However, the majority of such arguments maintain mutual independence of shocks, which rules out coheteroskedasticity or factor structures in volatility, an important feature of macroeconomic data (e.g., **?**) that TVV-ID can capture. **?** and **?** relax mutual independence to orthogonality, but the former paper introduces a cokurtosis assumption that still rules out correlated volatility processes, while the latter finds that estimators imposing this assumption have superior performance, and uses it for empirical work.

A final strength of TVV-ID is the testability of the identification conditions, as demonstrated in Theorem 3. This test exploits evidence of the identification conditions that can be found in the reduced form moment $E\left[\zeta_t \zeta_{t-p}'\right]$ directly. Conditions for identification via heteroskedasticity apply to otherwise unidentified structural parameters, generally making

[13]**?** offer an additional (local) identification argument for the GARCH model based on reduced-form moments, more similar to the TVV-ID argument.

[14]As a simple example, consider a univariate process with $\sigma_t^2 = 1, t = 1, \ldots, T/2$ and $\sigma_t^2 = 2, t = T/2 + 1, \ldots, T$. $cov\left(\sigma_t^2, \sigma_{t-1}^2\right) = \frac{1}{T-1}\left(2 + 1 \times (T/2 - 1) + 4 \times (T/2 - 1)\right) - \left(\frac{1+2}{2}\right)^2$, which converges to $2.5 - 1.5^2 = 0.25$ as T goes to infinity, so even with a single regime switch, the autocovariance is non-zero asymptotically.

testing difficult. This leads **?** to propose a test of the properties of some orthogonal rotation of the structural shocks, as noted above. In the regime setting, **?** introduce a novel testing approach that is also based on only reduced-form moments, and **?** develop a Bayesian procedure. Both papers find empirical evidence that could appear to question the non-proportionality of variances across regimes in various settings. Nevertheless, I successfully test for identification in my empirical setting. To assess the broader practical relevance of my approach, I also apply my identification test for TVV-ID's conditions to the two empirical applications in **?**, and an additional application from a 2018 working paper version. For the first two applications, my findings mirror those for the authors' test: the **?** model is identified, but that of **?** can only be partially identified based on time-varying volatility. For the third application, an extension of **?** found in **?**, the authors find only partial identification. In contrast, I easily reject all ranks less than n at the 1% level, establishing identification. These results suggest that my identification conditions (and associated test) may be at least as useful in practice as a regime-based scheme. Moreover, the latter finding lends support to the argument made above that the proportionality conditions required for TVV-ID may hold more generally than those for regime-based identification due to intra-regime dynamics.

3 Estimation and inference

In this section, I provide guidance to empirical researchers to apply TVV-ID in practice. I discuss reduced form estimation, estimators available to implement TVV-ID, inference approaches, and the remaining task of ordering the columns of H. I conclude with a step-by-step guide for the applied econometrician.

3.1 Estimation

Reduced form

TVV-ID may be applied to any data satisfying Assumptions A, B, and C. Typically this type of identification problem arises in innovations to a system of equations. The most common use case is SVARs. This means that a set of reduced-form parameters must be estimated in order to recover the innovations, η_t, as residuals. For example,

$$Y_t = a + \sum_{l=1}^{\bar{l}} A_l Y_{t-l} + \eta_t. \tag{8}$$

In my empirical application, Y_t is a 3×1 vector consisting of quarterly observations of (real) federal tax revenue, federal government consumption and investment, and GDP, with $\bar{l} = 4$. The reduced form parameters $\mathbf{A} = (a, A_1, \ldots, A_{\bar{l}})$ are estimated via OLS, providing estimated innovations $\hat{\eta}_t$ to recover H. I take $\hat{\mathbf{A}}$ and thus $\hat{\eta}_t$ as given in estimating H and assume that the shocks follow a symmetric distribution, which guarantees that the distributions of the reduced form and structural parameters are asymptotically independent (as in standard treatments, e.g., **??**). I maintain this assumption since it is consistent with the vast majority of volatility models proposed for identification via heteroskedasticity, and the preferred estimator based on my simulation study, discussed below, which I adopt in my empirical analysis.[15] Since the estimates are asymptotically independent, a researcher simply needs to use a (separately) valid asymptotic variance for each set of parameters.

Estimating H

Theorems 1 and 2 and Corollary 2 identify H up to column order. Before introducing estimators for H, it is important to note the relationship between a raw estimate $\hat{H}$ and the true parameter H (with unit diagonal and columns ordered to be consistent with some shock labeling). For standard asymptotic results to apply for an estimator $\hat{H}$, the estimator should explicitly select a certain matrix from the identified set of permutations and renormalizations of H. If the selection follows a statistical rule, like choosing the permutation and normalization that closest matches the identity matrix, or a lower triangular matrix, as in **?**, this step does not meaningfully restrict the identified set of matrices; **?** provide a detailed discussion. However, such a selection means that a consistent estimator $\hat{H}$ will in general be consistent for some *permutation and renormalization of* H, not H itself. I thus refer to a raw estimate of the columns of H as $\hat{H}$, with consistency implying $\hat{H} \xrightarrow{p} H_P$, where H_P represents some permutation of H. [16] In particular, I write

$$H_P P^* \times id\left(H_P P^*\right) = H, \tag{9}$$

[15]Without this assumption, while a two-step procedure remains valid, the covariance of reduced form and structural estimates must be computed for inference on IRFs, as discussed in **?**. I additionally show in Section 5 of the Supplement that a symmetry-like assumption is adequate for asymptotic independence to further hold for non-parametric estimators implementing TVV-ID. While it is also true that Generalized Least Squares estimating the reduced form and structural parameters jointly (or using a 2-step procedure as in **?**, for example) can deliver efficiency gains in the presence of heteroskedasticity, I do not pursue it here. Doing so introduces additional computational challenges contrary to my aim to make a simple proposal for practitioners, while complicating robustness checks comparing results across multiple implementations of TVV-ID, since each estimator begins with different innovations.

[16]If the selection rule incorporated in the estimator coincides with the labeling criterion adopted to interpret the shocks, as discussed in Section 3.2, then this preliminary selection step and the shock labeling step coincide, and $\hat{H}$ will be consistent for H.

where P^* is a permutation matrix and $id(HP^*)$ maintains the unit diagonal. I denote a labeled (permuted and renormalized) version of the estimator $\hat{H}$ as $\hat{H}_{lab}$.

Since TVV-ID is a non-parametric argument, it can justify both estimators making no parametric assumptions on the shock variance process and also estimators making a wide array of parametric assumptions on the variance process. In contrast, the previous identification approaches using heteroskedasticity make parametric arguments, which dictate very particular estimators. The former advantage of TVV-ID naturally points to estimating H via GMM, based on $cov\left(\zeta_t, \zeta'_{t-s}\right)$ and $E[\zeta_t]$. While fully non-parametric, this estimator faces challenges in small samples since it relies on imprecisely estimated higher moments for identification, features highly nonlinear estimation with a large parameter vector, and is overidentified (and thus sensitive to the choice of weighting matrix). For a sense of the challenge, the 3-variable SVAR in my empirical study leaves 27 parameters to be estimated from 42 equations; the 42 higher moments must be estimated from just 224 observations of Y_t. The latter advantage of TVV-ID points to a virtually limitless class of previously inadmissible parametric estimators for H, including those based on state space models, for example. It leaves the econometrician free to choose a model that she thinks suits an application well, and implement it without having to prove new identification results. Given this multiplicity of options, I provide guidance on applying TVV-ID in practice.

In the Supplement, I report a simulation study comparing various implementations of TVV-ID, as well as details on novel estimators, noting any additional assumptions required for consistency and asymptotic normality. I summarize the results here. I consider two estimators uniquely justified by TVV-ID: GMM estimates based on $cov\left(\zeta_t, \zeta'_{t-s}\right)$ and $E[\zeta_t]$ and an Expectation-Maximization (EM) estimator based on an AR(1) SV process for the shock variances, extending that of **?** and **?**. This state space model, adopted in **?** and **?**, for example, is described by

$$\begin{aligned} \varepsilon_t &\sim N(0, \Sigma_t) \\ \log \sigma_t^2 &= \mu(1-\phi) + diag(\phi) \log \sigma_{t-1}^2 + e_t,\ e_t \sim N(0, \Sigma_e). \end{aligned} \tag{10}$$

I additionally consider the GARCH(1,1) estimator of **?**, "Hybrid GARCH", which calibrates the autoregressive parameters, three Rigobon estimators (with regimes determined either by rolling windows as in **?**, an arbitrary $T/2$ sample split, or a Markov switching model as in **?**), and two estimators exploiting non-Gaussianity (FastICA of **?** and gJade of **?**, which accommodates possible stochastic volatility). These estimators have well-established asymptotic properties (consistency and asymptotic normality) for $\hat{H}$, and the interested reader should consult those references for details on limiting distributions.

I consider a range of heteroskedastic DGPs empirically calibrated to a bivariate system consisting of the Fed Funds rate and the first principle component of the remainder of the FRED-MD database. I compare mean estimates, RMSE, and the size of tests on true parameter values. I find that for all DGPs, the AR(1) SV model either performs best or very close to the leading estimator, even when badly misspecified. The Hybrid GARCH and Rigobon estimator based on a Markov switching model follow, with the non-Gaussianity estimators closely behind. The performance of GARCH and the remaining Rigobon estimators varies dramatically across DGPs. Finally, the fully non-parametric GMM approach does not perform well due to the difficulty in estimating fourth moments precisely in small samples. These results generally agree with related studies in **?** and **?**.

Accordingly, the AR(1) SV estimator is my preferred approach. I provide extensive details on this estimator in Section 2 of the Supplement, including asymptotic properties; I suggest using the inverse expected Fisher Information for inference. I provide ready-to-use code to implement the estimator as part of the replication files. In the empirical application, I focus on this preferred estimator. Given a lack of information furnishing obvious discrete regimes for fiscal policy, or evidence supporting particular functional forms in this data, the robustness to possible misspecification I find is appealing. However, I also consider sensitivity to estimators based on alternative variance processes, providing a blueprint for future empirical work to do the same. Alternatively, empirical researchers may also adopt model selection criteria as in **?** to select a suitable volatility model. Those authors do not consider the AR(1) SV model, while noting that the performance of their criteria depends on the true DGP and the model being tested; assessing the performance of model selection criteria for the AR(1) SV model is left for future work.

3.2 Labeling the structural shocks

To interpret the effects of the shocks ε_t or conduct inference for H, the ordered response matrix, the columns of $\hat{H}$ need to be ordered and renormalized. The researcher must now compare the implications of column permutations to economic information in order to determine the column labels and associated ordering that best correspond to how he or she has pre-defined the obejective, H (for example, stipulating that the first column contains the responses to a tax shock, the second to a spending shock). While this is often done informally in practice (e.g., **?**), I provide a more formal treatment, and show that the labeling indeterminacy does not impact inference on the labeled matrix $\hat{H}_{lab}$ asymptotically.

Let $f : \mathbb{R}^{n^2-n} \to \mathbb{R}^k$ be a $k-$vector-valued function of (permutations of) H that the researcher will compare to some $k-$vector of economic information, f_0. Let $P_i \in \mathcal{P}_n =$

$\{P_1, P_2, \ldots P_{n!}\}$ denote one of the set of $n!$ $n \times n$ permutation matrices. Finally, let $\rho : \mathbb{R}^k \to \mathbb{R}$ be some norm, for example the Euclidean norm. Any permutation and subsequent renormalization of H (or $\hat{H}$) can be written as $HP_i \times id\,(HP_i)$. Then define

$$\mathcal{L}_f\,(H, P_i, f_0) = \rho\,(f\,(HP_i \times id\,(HP_i))\,, f_0)$$

as a *labeling criterion*, with

$$\forall P_i \neq I_n,\ \mathcal{L}_f\,(H, I_n, f_0) < \mathcal{L}_f\,(H, P_i, f_0)\,. \tag{11}$$

$\mathcal{L}_f\,(H, P_i, f_0)$ compares the function f of $HP_i \times id\,(HP_i)$, a permutation of H, to the vector of economic information supplied by the researcher, f_0, under the norm ρ. Given H, $P_i = I_n$ is the unique minimizer of $\mathcal{L}_f\,(\cdot, \cdot, f_0)$.[17] It follows from the definition of H_P and P^* in (9) that $\mathcal{L}_f\,(H_P, P^*, f_0) = \mathcal{L}_f\,(H, I_n, f_0)$. With the addition of (11), H is no longer just identified up to column order, but is now globally point identified. There are many possible choices for f corresponding to intuitive labeling criteria. $\mathcal{L}_f$ can simply compare each element of H to its value under some prior, a previous study, or a structural model. It could do the same for an IRF function based on H. It chooses the permutation that minimizes the distance to f_0. Section 5 of the Supplement describes an extensive list of such labeling approaches. In my empirical study, since I compare my results to those of **?** and **?**, I simply use $f\left(\hat{H}P_i \times id\left(\hat{H}P_i\right)\right) = vec\left(\hat{H}P_i \times id\left(\hat{H}P_i\right)\right)$ and $f_0 = vec\,(H_{MR})$ (or $vec\,(H_{BP})$), so I choose the permutation that makes my estimates closest to those of the previous studies. This permutation is also *least conducive* to rejecting their results, rendering my analysis potentially conservative.

Define the permutation minimizing $\mathcal{L}_f\left(\hat{H}, P_i, f_0\right)$ as

$$\hat{P} = \underset{P_i \in \mathcal{P}_n}{\operatorname{argmin}} \mathcal{L}_f\left(\hat{H}, P_i, f_0\right),$$

and $\hat{H}_{lab}$, the labeled and unit-diagonalized, point identified estimate of H, as $\hat{H}_{lab} = \hat{H}\hat{P} \times id\left(\hat{H}\hat{P}\right)$. Theorem 4 outlines the asymptotic implications of labeling based on $\mathcal{L}_f\left(\hat{H}, P_i, f_0\right)$, and adopting the associated reordering and renormalization of columns, for inference on $\hat{H}_{lab}$:

Theorem 4. *If f is continuous and $\hat{H}$ is consistent for some permutation of H, H_P, and*

[17] In the process of specifying f, economic labels are associated with a particular column/shock ordering; for example, if f computes a function of $\hat{H}^{(1)}$ or $\hat{\varepsilon}_{1t}$ to compare to economic information on the behaviour of a tax shock, that implies that the first column of the objective matrix, H, is associated with the tax shock. Such an ordering may be natural if tax revenue enters the VAR first, rendering the unit diagonal normalization of H a named-shock unit effect normalization.

asymptotically normal, then $\mathcal{L}_f\left(\hat{H}, P_i, f_0\right)$ *is consistent for* $\mathcal{L}_f\left(H_P, P_i, f_0\right)$ *,* $\hat{P}$ *is consistent for* P^**, and* $\hat{H}_{lab}$ *is consistent for* H*; the labeling step does not impact the asymptotic distribution of* $\hat{H}_{lab}$ *beyond permutation and normalization.*

This result is based on **?**, who establishes asymptotic distributions in a discrete model selection setting, building on intuition dating to **?**.[18]

In practice, Theorem 4 greatly simplifies the task facing the econometrician. To conduct inference on $\hat{H}_{lab}$, the econometrician can simply either permute and renormalize the estimated structural parameters and then recompute the asymptotic variance directly (by reevaluating the moment covariance and Jacobian in the case of GMM, for example), or use the delta method to compute the asymptotic variance based on the asymptotic variance of $\hat{H}$ and the function $\hat{H}_{lab}{=}\hat{H}\hat{P} \times id\left(\hat{H}\hat{P}\right)$.

3.3 Framework for empirical analysis

The estimation considerations above suggest a series of straightforward steps that can be applied in practice to implement TVV-ID. The framework proceeds as follows:

1. Estimate the reduced form VAR as usual and obtain estimates of η_t.

2. Conduct a Cragg-Donald test of the rank of $\frac{1}{T-1}\sum vech\left(\hat{\eta}_t\hat{\eta}_t'\right) vech\left(\hat{\eta}_{t-1}\hat{\eta}_{t-1}'\right)'$ for $r = 1, \ldots n-1$ to (conservatively) verify that the model is identified.

3. Apply the selected (consistent and asymptotically normal) estimator of H exploiting persistent time-varying volatility to $\hat{\eta}_t$, taking $\hat{\eta}_t$ as given (my results recommend the AR(1) SV estimator, with replication code provided).

4. Use a labeling criterion $\mathcal{L}_f\left(\cdot, \cdot, f_0\right)$ satisfying the assumptions of Theorem 4 to obtain $\hat{P}$ and permute and renormalize $\hat{H}$ to obtain $\hat{H}_{lab}$; for inference, use permuted and renormalized estimates to construct the asymptotic variance of $\hat{H}_{lab}$ directly or apply the delta method to $\hat{H}$.[19]

[18] While I do not pursue it here, it is straightforward to extend these results to discontinuous $f\left(\cdot\right)$ (like several discussed in Section 5 of the Supplement), allowing for criteria based on rankings of elements of H by magnitude, for example.

[19] In particular, the delta method shows that $\sqrt{T}\left(vec\left(\hat{H}_{lab}\right) - vec\left(H\right)\right) \overset{d}{\to} N\left(0, \frac{\partial vec\left(\hat{H}_{lab}\right)}{\partial vec\left(\hat{H}\right)'} V_{H_P} \frac{\partial vec\left(\hat{H}_{lab}\right)'}{\partial vec\left(\hat{H}\right)}\right)$, where V_{H_P} is the asymptotic variance of $\hat{H}$, $\hat{H}_{lab}{=}\hat{H}\hat{P} \times id\left(\hat{H}\hat{P}\right)$, and the Jacobian of $vec\left(\hat{H}_{lab}\right)$ is easily computed analytically given $\hat{P}$.

5. Construct impulse responses and their asymptotic variance using an asymptotically valid variance for the reduced form parameters estimated in step 1 (I use the wild bootstrap of **?**), the asymptotic variance from step 4 ($V_{H_{lab}}$), and the delta method.[20]

I follow these steps in my empirical study, using the AR(1) SV estimator as my baseline.

4 New insights on fiscal multipliers from higher moments

The size of fiscal multipliers is central to policy debates, yet there remains considerable disagreement over their scale. The range of estimates is documented by **?**, **?**, and **?**, with influential studies finding multipliers ranging from 1 to 3 for both tax and spending. I use TVV-ID to obtain new estimates of both fiscal multipliers. I estimate a tax automatic stabilizer effect, a key parameter in the literature, of 1.58, closely aligned with estimates based on institutional details. I find dynamic multipliers peaking at 0.86 for tax cuts and 0.75 for government spending, while cumulative multipliers reach 2.06 and 0.87 respectively after 5 years. Accordingly, tax shocks explain up to 22% of variation in output after 2 years, while spending shocks have a smaller role. I show that the parameters identified by TVV-ID, and the higher moments that determine them, are consistent with our theoretical understanding of the relationship between fiscal shocks and output. In contrast, the popular methodology of **?**, which leads to much higher multipliers, implies surprising conclusions about higher moments, and faces concerns over instrument validity. Finally, the episodes identified by TVV-ID as most important accord well with the narrative record of fiscal policy and output in the United States.

4.1 Data & model

I apply TVV-ID to the trivariate VAR adopted by **?** (henceforth BP) and **?** (henceforth MR). The former paper forms a seminal baseline, while the latter attempts to reconcile

[20]Explicitly, for the $h-$horizon IRF Ψ^h, $h \geq 1$,

$$\sqrt{T}\left(\hat{\Psi}_{ij}^h - \Psi_{ij}^h\right) \xrightarrow{d} N\left(0, \frac{\partial \Psi_{ij}^h}{\partial\left(vec\left(R^h\right)', vec\left(H\right)'\right)} \begin{bmatrix} V_R & 0 \\ 0 & V_{H_{lab}} \end{bmatrix} \frac{\partial \Psi_{ij}^h}{\partial\left(vec\left(R^h\right)', vec\left(H\right)'\right)'}\right),$$

where $\Psi^h = R^h H$, R^h is the reduced form IRF ($R^h = \sum_{g=1}^h R^{h-g} A_g$, setting $R^0 = I_n$ and $A_g = 0, g > \bar{l}$), and $\frac{\partial \Psi_{ij}^h}{\partial\left(vec(R^h)', vec(H)'\right)} = \begin{bmatrix} H^{(j)\prime} \frac{\partial R_{(i)}^{h\prime}}{\partial vec(R^h)'} & R_{(i)}^h \frac{\partial H^{(j)}}{\partial vec(H)'} \end{bmatrix}$. **?** derive a valid heteroskedasticity-robust asymptotic variance in Theorem 2.1 for the reduced form parameters, which can be combined with equation 3.7.5 of **?** to estimate V_R. They also argue that the wild bootstrap of **?** is a valid estimator for V_R in the presence of conditional heteroskedasticity, and I adopt this procedure in my empirical analysis. Their Theorem 2.1 additionally provides expressions that hold without my assumption of symmetrically distributed shocks.

the existing range of fiscal multiplier estimates, including BP, with results obtained using the "proxy-SVAR" methodology of **?**, through the lens of this common model. These two influential papers form an ongoing basis of comparison for my results. The VAR consists of federal tax revenue, federal government consumption and investment, and GDP, based on quarterly BLS data found in the NIPA tables, spanning 1950 Q1 to 2006 Q4.[21] Additional details can be found in MR. I use the replication code available on Mertens' website to obtain identical reduced form results, with analysis then following the steps outlined in Section 3.3.

Turning to the structural identification problem, the baseline model is

$$\begin{aligned}
\eta_{Tt} &= \varepsilon_{Tt} + H_{12}\varepsilon_{Gt} + H_{13}\varepsilon_{Yt}, \\
\eta_{Gt} &= H_{21}\varepsilon_{Tt} + \varepsilon_{Gt} + H_{23}\varepsilon_{Yt}, \\
\eta_{Yt} &= H_{31}\varepsilon_{Tt} + H_{32}\varepsilon_{Gt} + \varepsilon_{Yt},
\end{aligned} \tag{12}$$

where $\eta_{Tt}, \eta_{Gt}, \eta_{Yt}$ are innovations to tax revenue, government spending, and output, respectively, and $\varepsilon_{Tt}, \varepsilon_{Gt}, \varepsilon_{Yt}$ are similarly labeled structural shocks. BP's structural model, using the notation of MR, can be written as a transformation of (12). Equation (13) states that model, where e_t are standardized structural shocks (so $E\left[e_t e_t'\right] = I$), with the mapping between their parameters and H (replacing MR's ζ_T with ξ_T to avoid confusion with $\zeta_t = vech\left(\eta_t \eta_t'\right)$):

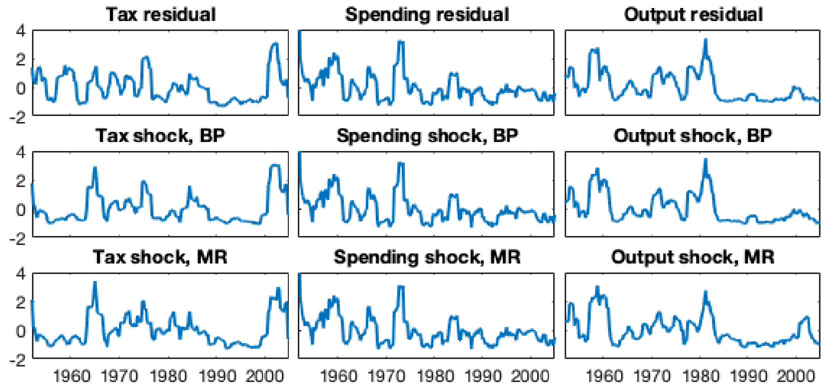
$$\begin{aligned}
\eta_{Tt} &= \sigma_T e_{Tt} + \theta_G \sigma_G e_{Gt} + \theta_Y \eta_{Yt}, & \theta_G &= \tfrac{H_{12} - H_{32}H_{13}}{1 - H_{23}H_{32}}, & \theta_Y &= H_{13}, \\
\eta_{Gt} &= \gamma_T \sigma_T e_{Tt} + \sigma_G e_{Gt} + \gamma_Y \eta_{Yt}, & \gamma_T &= \tfrac{H_{21} - H_{23}H_{31}}{1 - H_{31}H_{13}}, & \gamma_Y &= H_{23}, \\
\eta_{Yt} &= \xi_T \eta_{Tt} + \xi_G \eta_{Gt} + \sigma_Y e_{Yt}, & \xi_T &= \tfrac{H_{31} - H_{32}H_{21}}{1 - H_{21}H_{12}}, & \xi_G &= \tfrac{H_{32} - H_{31}H_{12}}{1 - H_{21}H_{12}}.
\end{aligned} \tag{13}$$

This mapping allows direct comparison of BP and MR parameters with the TVV-ID results. BP identify the model by calibrating the automatic stabilizer response of tax revenue, θ_Y, based on institutional data, and restricting $\gamma_T = \gamma_Y = 0$. MR instead estimate γ_T and ξ_T using the **?** (henceforth RR) tax shocks as a proxy variable, and restricting $\gamma_Y = 0$ to separately identify spending and output shocks.

I first assess whether the data supports TVV-ID in this setting. Figure 2 plots evidence of heteroskedasticity in the data, moving averages of the squares of reduced form residuals, BP shocks, and MR shocks in turn; in all series, there appear to be strong patterns of heteroskedasticity. Next, I formally test the rank identification condition, exploiting Theorem 3. As I discuss in Section 4.3, this condition will hold if the variances of certain shocks have

[21] While **?**, for example, estimate a TVP model in the reduced form, I maintain the constant parameter reduced form VAR specification, since that remains the benchmark in the literature, even in more recent work, such as **?** and **?**. Maintaining the same reduced form allows a clearer comparison of identifying assumptions across approaches.

Figure 2: Moving averages of squared residuals and shocks



2-year moving averages of the square of the specified series (standardized).

limited predictive power for the variances of others, for example. I test the null hypothesis that the rank, r, of $\frac{1}{T-1}\sum \hat{\zeta}_t \hat{\zeta}_{t-1}'$ is 1 (against the alternative $r > 1$), then the null hypothesis $r = 2$ (against $r > 2$). In this 3-variable system, a rank of 3 implies that $\begin{bmatrix} M_{t,t-1} & E\left[\sigma_t^2\right] \end{bmatrix}$ satisfies the conditions of Theorem 2. The tests, reported in Table 4 in the Supplement, reject the null ranks at the 10% and 5% levels respectively, jointly indicating a rank exceeding 2, so the conditions for TVV-ID are satisfied. This conclusion is unchanged even after accounting for multiple testing using a Bonferroni-Holm correction, under which these tests jointly reject at the 10% level.[22]

4.2 Results

In this section, I present the headline empirical estimates based on TVV-ID. I first show that the values of two key structural parameters differ from those of BP and MR. Accordingly, TVV-ID estimates a considerably lower dynamic multiplier for tax cuts, while the spending multiplier remains similar; both are below 1. The same is true for cumulative multipliers, based on which it becomes clear that tax cuts do, nevertheless, have stronger effects on output with multipliers around 2 after 20 quarters. Finally, I show that tax cuts explain a greater share of variation in output through forecast error variance decompositions, reaching 22% after 8 quarters.

I first present estimates for the structural parameters in equation (13). These are simple transformations of $\hat{H}_{lab}$, which I estimate using the AR(1) SV model discussed in Section 3, with additional details in Section 2 of the Supplement. TVV-ID requires the estimated shocks to be labeled; I do so using a labeling criterion that chooses the permutation of $\hat{H}$

[22]The Bonferroni-Holm $p-$values for a family of two hypotheses at the 10% level are 0.10 and 0.05. The test of $r_{null} = 1$ meets the first and the test of $r_{null} = 2$ meets the second.

Table 1: Estimates of structural parameters

	θ_G	θ_Y	γ_T	γ_Y	ξ_T	ξ_G
TVV-ID	-0.13	1.58^{***}	0.11	0.02	-0.00	0.06
	(0.10)	(0.18)	(0.13)	(0.39)	(0.02)	(0.045)
BP	-0.06	2.08	0	0	-0.08	0.07
MR	-0.20	3.13	0.06	0	-0.35	0.10

The first row maps estimates of H obtained via TVV-ID to the parameters of BP and MR using (13). TVV-ID is implemented using the AR(1) SV model, described in Section 3, with details provided in Supplement Section 2. The remaining rows are estimates from **?**, for comparison.

that is the closest match to BP and MR.[23] Table 1 reports the results, with BP and MR estimates for comparison.[24] The automatic stabilizer effect for tax revenues, θ_Y, is estimated to be 1.58, and is statistically significant. The 95% confidence interval, $[1.23, 1.94]$, does not include the BP or MR value. While the estimate for the instantaneous response of output to tax revenues, ξ_T, is not statistically distinct from zero, it is estimated precisely enough to reject the negative values of both BP and MR at the 1% level. The remaining parameters are quite similar across specifications; I do not reject the zeros assumed by BP and MR as identifying restrictions. However, the two parameters for which I do obtain distinct values represent the novel identifying information in BP and MR: BP externally calibrate the automatic stabilizers, θ_Y, while MR estimate ξ_T using an external instrument. **?** show that these two parameters are crucial for determining fiscal multipliers. Higher values of θ_Y mechanically dictate higher multipliers by implying negative values for ξ_T to match reduced form covariances. Moreover, ξ_T, the contemporaneous response of output to tax revenues, directly implies the size of the tax multiplier.

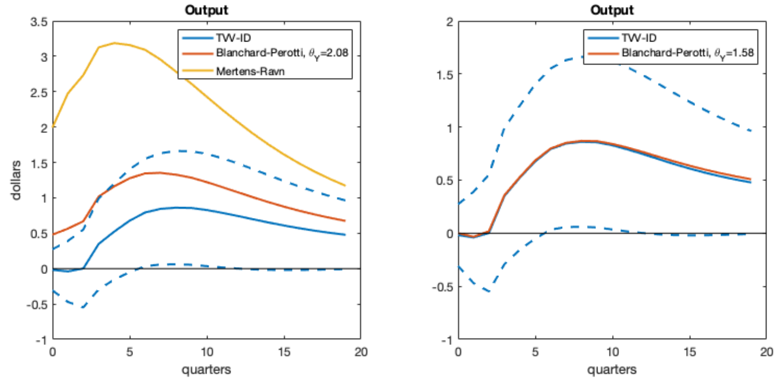
One might be concerned that the validity of statistical inference is impacted by the use of pre-tests for identification. However, given the $p-$values of the identification tests in this application, it turns out that any statistical test that rejects at the 5% level – like that of the BP or MR values for θ_Y – can be jointly rejected with the identification pre-tests at the 10% level, using the Bonferroni-Holm adjustment.[25] All subsequent results that are significant at the 5% level can also be interpreted accordingly.

[23] In this application, this is the one clear labeling. As a further check, Figure 10 in the Supplement plots dynamic multipliers for all alternative labelings; it is clear that the labeling selected is the only one that produces results of plausible sign and magnitude.

[24] It is well-known that EM algorithms can be sensitive to start values; thus, optimization was carried out across a grid of start values and the median estimates were used to initialize a final optimization. The range of estimates across start values is very small, see Table 7 in the Supplement. As an additional check, the estimates from alternative volatility models (same Table) are quite similar.

[25] For a family of three hypotheses, the Bonferroni-Holm $p-$values for a 10% test are 0.10, 0.05, and 0.03. The test of $r_{null} = 1$ meets the first $p-$value, and the test of $r_{null} = 2$ meets the third. Thus, for a joint rejection at the 10% level, any additional test must have a $p-$value of 0.05 or smaller.

Figure 3: Response to a $1 tax cut



Dashed lines are 95% confidence intervals. The BP estimates in the left panel use their elasticity $\theta_Y = 2.08$; the right uses the value of 1.58 estimated via TVV-ID.

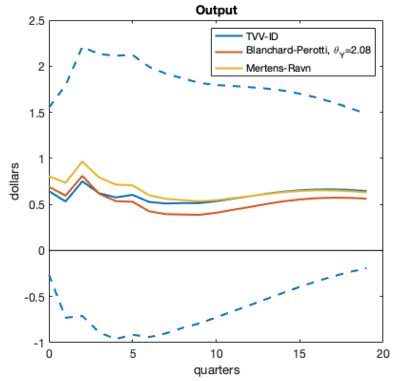
The implications of the estimates for θ_Y and ξ_T are evident in the dynamic multipliers. The left panel of Figure 3 plots the dynamic multiplier to a $1 tax cut based on TVV-ID (in blue), with a 95% confidence interval.[26] The responses obtained using the BP and MR approaches are plotted for reference (red and gold). The point estimate of the response of output to a tax cut remains approximately zero for the first two quarters, before rising. It peaks at 0.86 after 8 quarters. In contrast, BP and MR obtain responses on impact of 0.48 and 1.99 respectively, with peak effects of 1.35 (7 quarters) and 3.19 (4 quarters) respectively. The near-zero estimated value of ξ_T under TVV-ID means that there is a response lag for output, with effects coming through the lag structure of the VAR. The very large multiplier of MR is rejected at all horizons, while the BP response is only rejected through 2 quarters. The right panel plots the BP response, with the calibrated θ_Y replaced by the value estimated via TVV-ID. With this alteration, the responses are essentially identical. This finding bears out the argument of **?** that θ_Y essentially pins down the multiplier in this model.

Figure 4 plots the government spending multiplier. The multiplier is 0.65 on impact, peaks at 0.75 after 2 quarters, and is quite persistent, although the response is imprecisely estimated. Recall that the parameters linking spending and output in Table 1 are very similar across models; accordingly, so too are the multipliers. BP and MR yield 0.69 and 0.80 on impact, respectively, and 0.81 and 0.96 after two quarters. Figure 13 in the Supplement reports both dynamic multipliers for subsamples of the data. The results are fairly robust to omitting periods featuring key episodes. Spending multipliers peak as high as 1.2 (1980-2000 subsample), while peak tax cut multipliers fall as low as 0.5 (1970-2000 subsample).

Since the shape of the dynamic multipliers differs for each policy measure, it is natural to

[26] As described in Section 3.3, I compute the confidence intervals based on a **?** wild bootstrap variance for the reduced form parameters, as in **?**, the methods described in Section 2 of the Supplement for the structural parameters, and the delta method.

Figure 4: Response to a $1 spending shock



Dashed lines are 95% confidence intervals.

compare cumulative multipliers. I do so using the present-value formula of **?**; this approach also accounts for a persistent response in either tax revenue or government spending in the denominator of the multiplier. Table 2 reports the results with BP and MR for comparison. As for the dynamic multiplier, it takes some time for tax shocks to have an effect, but the response reaches 0.71 after 2 years, and 2.06 after 5 years. Again, these responses are considerably smaller than those based on BP and MR. Turning to spending, the cumulative multiplier is relatively consistent across horizons: 0.65 on impact, 0.57 after 2 years, and 0.87 after 5 years. These responses are similar to BP and MR. The results reaffirm that while the effects of a tax cut are not immediate, at longer horizons they are larger than those for spending. These results favour models where real activity responds to tax shocks with a delay. Table 10 in the Supplement shows that the cumulative multiplier results are robust for subsamples and alternative estimators based on time-varying volatility.

My results align with the literature more broadly. Beginning with the structural parameter θ_Y, TVV-ID matches estimates of automatic stabilizers based on institutional details of tax revenues remarkably well. In particular, **?** estimate the elasticity of tax revenues with respect to output for the federal government, and obtain a value of 1.6 for the period 1986-2008 – nearly identical to the value of 1.58 I obtain via TVV-ID. They estimate 1.4 for 1960-1985. In their calibration, BP use the value 2.08, but crucially compute this number based on data on *general* government revenue, which will naturally be higher than that for just the federal government, the data used here. As shown in Figure 3, the discrepancy between TVV-ID and BP can be entirely explained by using a value of θ_Y that is tailored to the *federal* government data used in the VAR.

Turning to the dynamic multipliers, my estimates accord well with those of **?**, who use *non-fiscal* proxies as external instruments. Figure 11 in the Supplement plots their estimates against mine; in general, the IRFs are similar, with their responses lying within my 95%

Table 2: Present value cumulative multipliers

	Tax						
	Impact	2Q	4Q	8Q	12Q	16Q	20Q
TVV-ID	-0.02	-0.03	0.09	0.71	1.33	1.77	2.06
Blanchard & Perotti	0.48	0.64	1.01	2.28	3.63	4.69	5.41
Mertens & Ravn	1.99	3.25	6.21	–	–	–	–
	Spending						
	Impact	2Q	4Q	8Q	12Q	16Q	20Q
TVV-ID	0.65	0.56	0.57	0.57	0.64	0.76	0.87
Blanchard & Perotti	0.69	0.61	0.60	0.56	0.60	0.70	0.80
Mertens & Ravn	0.80	0.73	0.73	0.71	0.78	0.89	1.00

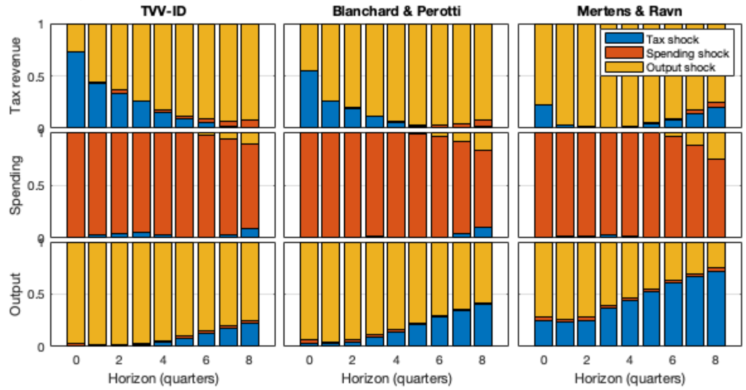
Present value cumulative multipliers computed using the IRFs reported above and the formula in **?**, with the average effective Fed Funds rate over the sample used as the discount rate. I do not report multipliers for horizons where the fiscal response in the denominator crosses zero, causing explosive behaviour.

confidence interval, and mine within their 95% credible set. When they calibrate a version of BP using $\theta_Y = 1.7$, motivated by **?**, they obtain very similar responses to mine.

The cumulative spending multipliers obtained also closely match results of **?** based on defense-related spending events. Her estimates range from 0.6-0.8, as do mine. While **?** obtain higher estimates for some states of the economy using both defense spending and BP spending shocks as instruments, the 0.6-0.8 range generally accords with their results as well. On the other hand, my cumulative multipliers differ from those of **?**, who find much larger responses to tax shocks and a reversal in the sign of the spending response after 8 quarters. Indeed, **?** is one of few papers to obtain tax multipliers on the scale of MR. **?** show that the **?** value for θ_Y, about 3.2, is linked to those authors' penalty-function identification approach, which *maximizes* the systematic component of tax revenues.

Finally, I also characterize the ability of each shock to explain unforecasted variation in observable series. This leverages both the estimated dynamic multipliers and the second moments of the shocks implied by my identification approach. Forecast error variance decompositions (FEVDs) compute the share of the $h-$step ahead forecast error variance accounted for by each structural shock. I plot FEVDs for the first 8 quarters in Figure 5. Each row corresponds to a different observable series, and each column to an alternative identification approach. TVV-ID finds that at short horizons most variation in tax revenue is driven by tax shocks, with the automatic stabilizer effect naturally accounting for a larger share at longer horizons. At all horizons, virtually all variation in spending comes from spending shocks. At short horizons, nearly all variation in output is explained by output shocks (which in this simple model likely encompass many more-specific shocks), but at longer horizons, the

Figure 5: Forecast error variance decompositions



Forecast error variance decompositions computed based on the IRFs for each identification approach.

share attributable to tax shocks rises to 22%; in contrast, spending shocks explain at most about 3%. These findings further illustrate that tax shocks may trigger stronger responses in output. The BP results almost exactly replicate those of TVV-ID, as do those of MR for spending. However, the MR results imply that the majority of variation in tax revenue at all horizons is explained by output shocks; this is not implausible, and simply hinges on the alternative schemes disagreeing on the relative importance of automatic stabilizers versus policy changes. More puzzlingly though, the decomposition ascribed by MR implies that even over the first quarters, tax shocks explain 25% of variation in output, with that share rising to about 75% by 8 quarters. This story, of tax shocks having such immediate effects on output, and ultimately being the *most important* determinant of output fluctuations, is at odds with conventional wisdom and extensive prior work (finding no such role for tax shocks, as in **?**, discussed in detail below, or suggesting, for example, that financial or investment conditions are the main driver of business cycles, e.g., **????**). Through the FEVDs, the second moments implied by the alternative identification schemes are of policy relevance, since they imply the relative ability of fiscal shocks to explain output fluctuations. I find that tax shocks have greater potential to shape output fluctuations than spending shocks. I further illustrate how higher moments can be informative for identification in the next section.

4.3 Interpreting the identifying variation

Having described the response of output to fiscal shocks and highlighted discrepancies with the previous results of MR (and to a far lesser extent, BP), I now characterize the identifying variation exploited by the approaches. I argue that the variation exploited by TVV-ID is not just statistical in nature, but *economically* meaningful. In particular, what one might conclude about the relationship between output and tax shocks, based on their higher moments,

Table 3: Unconditional autocorrelation of structural shock variances

	TVV-ID			Blanchard & Perotti			Mertens & Ravn		
	$\sigma^2_{T,t-1}$	$\sigma^2_{G,t-1}$	$\sigma^2_{Y,t-1}$	$\sigma^2_{T,t-1}$	$\sigma^2_{G,t-1}$	$\sigma^2_{Y,t-1}$	$\sigma^2_{T,t-1}$	$\sigma^2_{G,t-1}$	$\sigma^2_{Y,t-1}$
$\sigma^2_{T,t}$	0.87^{***}	0.40^{***}	-0.08	0.78^{***}	0.37^{***}	0.04	0.70^{***}	0.38^{***}	0.47^{***}
$\sigma^2_{G,t}$	0.48^{***}	0.99^{***}	0.66^{***}	0.51^{***}	0.98^{***}	0.70^{***}	0.55^{***}	0.99^{***}	0.93^{***}
$\sigma^2_{Y,t}$	-0.10	0.65^{***}	0.99^{***}	0.05	0.69^{***}	0.99^{***}	0.62^{***}	0.89^{***}	0.97^{***}

Autocorrelation computed using the filtered variance paths obtained via TVV-ID and filtered paths estimated by fitting a multivariate AR(1) SV model to the shocks implied by **?** and **?** structural shocks. Results are starred at the 1, 5, and 10% levels.

is informative for whether those shocks have been successfully identified when compared to theory. This analysis serves both to bolster the credibility of my findings, identified using these moments, as well as to offer novel empirical results. In contrast, I show that the external instrument exploited by MR could be invalid, since both relevance and exogeneity may be rejected.

TVV-ID identifies H from the autocovariance of the squared reduced form innovations, and thus, implicitly, as coefficients on the autocovariance of structural variances in equation (6). These autocovariances of structural variances are identified jointly with H; any rotation of H implies a rotation of the structural shocks, and thus different properties for the autocovariance of their variances. Table 3 reports the (unconditional) autocorrelations of these variances, based on filtered variance paths.[27] Estimates are starred at 1, 5, and 10% levels.

TVV-ID suggests that each of the shock variances is strongly persistent. There is also strong *co-persistence* exhibited between tax and spending variances and spending and output variances. However, there is no such co-persistence between tax shock and output shock variances. The results are strikingly similar for the BP shocks. However, the rotation of H associated with the MR identification approach additionally imposes a strong co-persistence between tax shock variances and output shock variances. What does this co-persistence imply? A positive correlation of $\sigma^2_{Y,t-1}$ and $\sigma^2_{T,t}$ suggests that periods of volatile structural shocks to output predict subsequent periods periods of unexpected tax changes. Such a relationship contradicts the intuition of the RR narrative shocks and the MR identification. The point of RR's series is to isolate purely exogeneous tax policies independent of business cycle movements. The positive correlation suggests that, on average, large tax movements identified by this instrument are in fact *predicted* by business cycle movements.

In the other direction, a positive correlation of $\sigma^2_{T,t-1}$ and $\sigma^2_{Y,t}$ suggests that periods of

[27] I report the correlations, not covariances, to render the scale more interpretable and comparable across series. For TVV-ID, the variances are obtained as a byproduct of the estimation via the EM algorithm. For MR and BP, I apply a version of the EM algorithm to the estimated shocks that treats the input series as already orthogonalized, and simply fits a multivariate AR(1) SV process.

unexpected tax changes predict subsequent periods of volatile structural shocks to output. Since these are structural shock variances, this correlation is not simply output moving endogenously following tax changes. In this case, economic theory can be helpful. **?** merge an incomplete markets model with a New Keynesian model to capture the U.S. tax and transfer system and show that changes to tax rates (which would be tax shocks in this empirical setting) have only negligible effects on aggregate volatility. This result holds for both the level and intercept of the tax schedule. The intuition is that while changes in tax rates alter labour supply or investment decisions, they do not do so differentially across the business cycle, and thus have little effect on volatility. This result is fully consistent with the zero co-persistence uncovered by TVV-ID and BP, while inconsistent with the strong positive relationship of MR. TVV-ID in turn provides empirical support for their structural model, by obtaining similar findings based on statistical properties of the macroeconomic data.

The facts detailed above are central in distinguishing between TVV-ID and the MR approach. The autocovariances implied by TVV-ID are more consistent with the higher moments theory predicts for tax and output shocks. They illustrate the information that TVV-ID, by virtue of selecting a rotation of the shocks that is additionally consistent with such higher moments of the data, leverages to rule out rotations like that of MR. The same is true for the FEVDs reported above and the paths of the shock variances discussed below.

The two "zero" autocovariances also provide an economic reason for why the rank condition for identification should hold. The matrix of autocorrelations in Table 3 is a rescaled subset of $M_{t,t-1}$, which Theorem 1 requires to have rank of at least 2, with no proportional rows. If each structural variance is persistent, then inserting these two zero co-persistences implies that the rows corresponding to $\sigma^2_{T,t}$ and $\sigma^2_{Y,t}$ are linearly independent (so rank is at least 2). Unless $\sigma^2_{G,t}$ is almost perfectly predicted by one of $\sigma^2_{T,t-1}$, $\sigma^2_{Y,t-1}$ but not at all by the other, the $\sigma^2_{G,t}$ row also cannot be proportional to either of these linearly independent rows. The same is true for $M_{t,t-1}$ as a whole, so identification holds. Thus, the seemingly abstract statistical conditions applied to $M_{t,t-1}$ can have clear economic interpretations.

In contrast to the higher moments of TVV-ID, MR use the RR shocks as external instruments to identify tax shocks. As discussed in **?**, like standard instrumental variables, such proxies must be both relevant and exogenous. I test the relevance condition using the first-stage $F-$statistic under the alternative assumptions of either homoskedasticity or heteroskedasticity, where the critical values are $F > 10$ (**?**) or $F > 23$ (**?**), respectively. The corresponding $F-$statistics are 4.13 and 1.76, so weak identification cannot be rejected.[28,29] It is also possible to test the exogeneity of MR's instrument. Under the assumption that the TVV-ID estimates are valid, I regress the RR proxy on $\hat{\varepsilon}_t$ and test the hypotheses that coefficients on ε_t^G and ε_t^Y are zero. The test rejects at the 5% level for the shocks jointly, driven by a significant negative relationship with ε_t^Y. After a Bonferroni-Holm adjustment to account for identification pre-tests, as discussed above, these tests still reject at the 10% level.[30] This test is conditional on the parameters identified via TVV-ID, and thus assumes that the identifying conditions for TVV-ID hold, in order to test the MR assumptions. However, these assumptions were verified; furthermore, the results are very similar using the BP shocks, suggesting they are robust to small changes in the estimated structural parameters. Table 8 in the Supplement repeats both the relevance and exogeneity tests for the alternative narrative measures considered by MR; the findings are robust. The rejection of exogeneity parallels the previous finding of correlation between the variance of output shocks and the variance of tax shocks. It appears that despite careful construction, the narrative proxy exhibits counter-cyclical behaviour. Given the need to parse political motivations behind tax changes in order to classify them in RR, it is possible that in an effort to avoid pro-cyclicality, the time series may over-omit ideologically-motivated events that might appear pro-cyclical. The strong negative relationship between the instrument and output shocks implies that, for a tax cut, the estimated impact on output is likely biased upwards, which is consistent with the large multipliers MR recover. **?** use a similar methodology, but with *non-fiscal* proxies that they confirm satisfy both relevance and exogeneity conditions, and find much smaller multipliers.

4.4 Structural shocks and the narrative record

The fiscal shocks that I recover also match the narrative record well and cast light on particularly important policy episodes. First, I describe how the analytical relationship between ξ_T and θ_Y can lead MR to differ in decomposing observed comovements into tax shocks or output shocks. I then characterize the most important tax episodes over the sample regard-

[28] These results are at odds with the reliability measure MR report. This measure is asymptotically equivalent to the R^2 associated with the instrument. It can only be computed based on estimated structural shocks; instrument validity is *assumed* to obtain these. The $F-$statistic is more informative since thresholds based on the bias of IV estimators are available. Additionally, the reliability statistic makes assumptions on the form of measurement error.

[29] While the instruments considered here appear weak, **?** focus on the impact of *marginal* tax rates and construct an alternative instrument based on the RR narrative shocks, scaled based on marginal tax rate changes, which appears to be a strong instrument for the tax rate changes they consider.

[30] Additionally, adjusting inference to account for $\hat{\varepsilon}_t$ being a generated regressor following **?** leaves test statistics changed by at most 0.01. These exogeneity test results are also surprisingly robust to alternative labelings of the tax shock, as shown in Table 9 of the Supplement.

less of identification approach, mostly reforms of the 1960s and 1970s and the Bush tax cuts. Next, I describe how, as a result of how they estimate ξ_T and θ_Y, TVV-ID and MR differ in their classification of key episodes as tax versus output shocks; most strikingly, TVV-ID characterizes the 1981-82 recession as mainly an output shock, while MR explains it using a tax shock. The same discrepancies appear when analyzing variance paths of the shocks.

Both BP and MR use external information to supply identifying restrictions to identify the model, with BP calibrating θ_Y and MR estimating ξ_T. As argued in the previous section, there is reason to doubt this estimate of ξ_T. In a bivariate version of the baseline model without spending, there is a simple analytical relationship between the parameters θ_Y and ξ_T. In particular, define the bivariate model as

$$\eta_t^{bi} = H^{bi}\varepsilon_t^{bi},\ E\left[\eta_t^{bi}\eta_t^{bi\prime}\right] = \Sigma^{bi} = \begin{bmatrix} \sigma_{\eta_T}^{2,bi} & \sigma^{bi} \\ \sigma^{bi} & \sigma_{\eta_Y}^{2,bi} \end{bmatrix}.$$

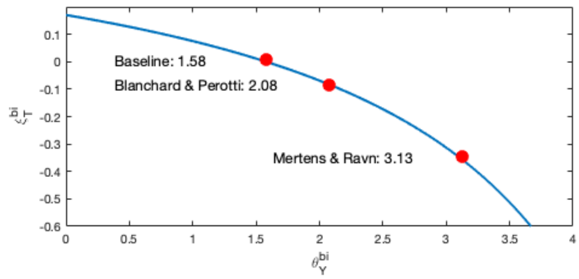
This essentially coincides with the "simple fiscal rule" case in **?**. The instantaneous effect of taxes on output is given by

$$\xi_T^{bi} = \frac{\theta_Y^{bi}\sigma_Y^{2,bi} - \sigma^{bi}}{\theta_Y^{bi}\sigma^{bi} - \sigma_T^{2,bi}}, \tag{14}$$

which, after normalization, is equivalent to the result in equation (11) in **?**. Given the empirical properties of η_t^{bi}, ξ_T^{bi} and θ_Y^{bi} are inversely related. Figure 6 plots the value of ξ_T^{bi} as a function of θ_Y^{bi}, with the point on the curve chosen by each identification scheme marked. While TVV-ID chooses a point near zero, with BP's θ_Y implying a mildly negative response, MR's estimate is strongly negative, -0.35. TVV-ID allows the data internal to the model, including the higher moments discussed above, to choose the point on this curve. However, both other approaches, by choosing a point externally, eliminate all but one possible rotation, and in the case of MR, this rotation may have undesirable implications. In particular, since it implies such a large positive comovement through the automatic stabilizer effect, it means that episodes featuring a negative comovement between tax and output will be mostly explained by the model as a tax shock, while TVV-ID will interpret them as some combination of a tax shock and an output shock. As discussed below, the latter appears more consistent with the narrative of the U.S. macroeconomy.

Turning to the narrative record, I compare how TVV-ID and MR interpret important episodes as tax and output shocks to assess which rotation of the structural shocks is more consistent with the narrative economic evidence. The spending series are virtually identical, so I defer discussion to the Supplement; I omit the BP series, since they are quite similar to TVV-ID. Section 1.2 of the Supplement plots the shocks and discusses in detail all periods with substantial discrepancies between the alternative tax and output shocks. I summarize

Figure 6: Automatic stabilizers and the tax multiplier



The curve plots ξ_T^{bi} as a function of θ_Y^{bi}, equation (14), using the BP parameters to compute reduced form moments absent spending shocks (other models imply virtually identical curves). The red circles indicate the value on the curve chosen by each identification scheme.

key findings here, which offer insights into the plausibility of the two sets of shocks.

Before discussing discrepancies though, there are some episodes that are important for identification, no matter the identification approach. There are 7 common shocks amongst the 10 largest magnitude shocks for both TVV-ID and MR. In chronological order, they are:

1951 Q1: Truman's Revenue Act of 1950

1964 Q2: Johnson's Revenue Act of 1964

1965 Q3: Continued effect of Revenue Act of 1964 or Excise Tax Reduction Act of 1965

1975 Q3: Ford's Tax Reduction Act of 1975

2001 Q3: Bush's Economic Growth and Tax Relief Reconciliation Act

2002 Q1: Bush's Job Creation and Worker Assistance Act of 2002

2003 Q3: Bush's Jobs and Growth Tax Relief Reconciliation Act

All of these shocks also correspond to important fluctuations in the variance paths of Figure 7, which constitute identifying variation for TVV-ID.[31] Three of them are among the 18 observations that drive the negative relationship between η_{Yt} and the instrument in MR's identification, including the second most influential such observation (2003 Q3).

There is one consistent theme throughout the periods of discrepancy considered. For certain key dates, events that TVV-ID interprets as negative output shocks register as positive tax shocks in MR, and vice versa. This is due to the fact that ξ_T is near-zero in TVV-ID, but strongly negative in MR. As noted above, this means that when tax revenues and output move in opposite directions, MR is likely to attribute the episode to a tax shock, whereas TVV-ID will explain it with a combination of opposite-signed tax and output shocks. The most serious case of this is 1981 Q2, the onset of the 1981-82 recession, a clear case for a negative output shock. However, MR puzzlingly attribute the recession to their third largest

[31] Except 1951 Q1, since, as the first observation, its variance is an initial condition in the EM algorithm.

tax increase. Conversely, TVV-ID registers its third most contractionary output shock, consistent with the macroeconomic narrative. In brief, the cases include:

1971 Q1: MR register their sixth largest tax cut, coinciding with only modest changes to depreciation rules, under conditions consistent with an expansionary output shock.

1974 Q3: MR register their fifth largest tax increase, with no recognized policy change, under conditions more consistent with a contractionary output shock.

1978 Q2: MR register their eighth largest tax cut, with no recognized policy change, to help explain a large output boom.

1981 Q2: MR register their third largest tax cut to explain the 1981-82 recession, while TVV-ID recovers its third most contractionary output shock.

1990 Q4: MR register a large tax increase, with no recognized policy change, to help replicate the trough of the 1990 recession.

2001 Q3: MR put the 9/11 output shock on par with the trough of the five-times-deeper 1980 recession in order to offset the effect of a contemporaneous tax cut.

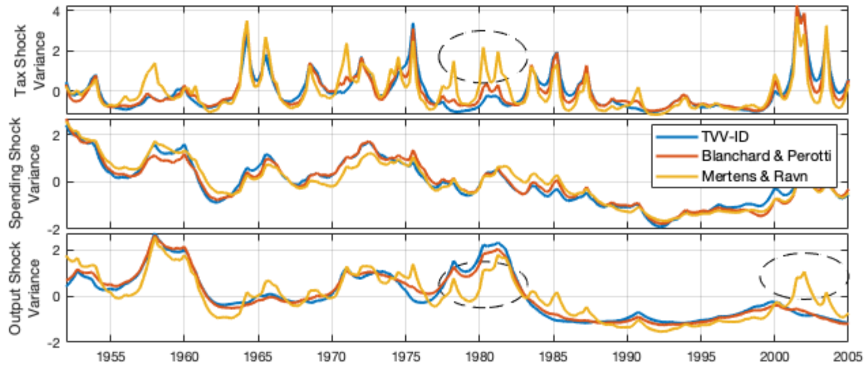
While certain anomalies also appear in the TVV-ID shocks, the majority of surprising results arise in the MR shocks, and can generally be linked to the substantially negative ξ_T.

Finally, I compare filtered paths of structural variances for each set of shocks, Figure 7, to the same narrative record. Since the AR(1) SV estimator fits variance paths to the data, these are auxiliary identifying moments for my estimate for H. The variance paths are generally similar across approaches, particularly TVV-ID and BP, with two notable exceptions for MR, which are circled on the plots. First, MR registers spikes in the variance of tax shocks in the late 1970s and early 1980s, commensurate with its lower variance in output shocks at those times. Based on the narrative above and additional events detailed in the Supplement, this further establishes the tendency of MR to interpret likely output shock variation as tax shock variation. Second, this rotation of shocks registers a high variance period for output shocks in the early 2000s, following the 2001 recession, nearly as pronounced as that of the early 1980s, a period widely recognized as considerably more volatile. Separately, these variance paths may be objects of economic interest; the variances for both tax and spending shocks can be seen as measures of respective policy uncertainty, while the output variance path traces well the familiar narrative of the Great Moderation.[32,33]

[32]Note, however, that the variance of forecast errors is not necessarily a natural analog to true "uncertainty", but is frequently discussed as such, see for example **?**.

[33]Interestingly, spending "uncertainty" peaks during the Korean War, various periods of the Vietnam war, and again with Afghanistan/Iraq.

Figure 7: Paths of structural shock variances



Standardized filtered variance paths estimated via TVV-ID or fitting an AR(1) SV model to ? and ? structural shocks. Dates of key discrepancies are circled.

5 Conclusion

This paper presents a general argument that structural shocks can be identified via time-varying volatility. The previous literature offers identification arguments based on a path of variances available for only a handful of parametric models of the variance process. My identification approach makes minimal assumptions on the variances as a stochastic process. This argument highlights a novel channel of identification based on heteroskedasticity that frees the researcher from needing to assume a particular functional form (or, indeed, any functional form) to obtain identifying moments. On the other hand, this also empowers researchers to develop new parametric models and approaches in contexts exhibiting time-varying volatility without needing to re-establish identification for each. Based on extensive simulations, one such novelly identified model, an AR(1) SV process, appears robust to misspecification and well-suited for applied work; fully non-parameteric estimators struggle in small samples. Economic information commonly used to structurally identify such models now need only be used to label the shocks identified by TVV-ID. I propose a simple test of the identification conditions based only on reduced form moments. I provide a step-by-step guide for empirical researchers to exploit TVV-ID.

I use my methodology to estimate fiscal multipliers. I estimate peak multipliers of 0.86 for tax cuts and 0.75 for spending. Cumulative multipliers reach 2.06 and 0.87 respectively after 20 quarters. I further find that tax shocks explain substantial variation in output at longer horizons, up to 22%. The spending multipliers reaffirm most previous results, including ?, ?, ?, and ?. The larger tax multipliers of ? can easily be reconciled with TVV-ID by adjusting their calibrated automatic stabilizer effect to the value I estimate based on TVV-ID. This value, about 1.6, is remarkably similar to the estimate of ?. I illustrate that the higher moments exploited by TVV-ID for identification have real economic content, and are consistent with theoretical work, including **?**. In contrast, the estimates of **?**, which imply a much larger tax multiplier, can be ruled out based on these moments. I also show that their instrument for fiscal shocks does not pass tests for validity. In addition, I highlight policy episodes that are key to identifying tax multipliers, and argue that the shocks implied by TVV-ID are more consistent with the narrative record than those of **?**. I thus provide new evidence, including higher moments, to reject some of the largest tax multipliers found in the literature, while offering support for the increasing consensus on multipliers closer to unity.

A Proofs

A.1 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 $E_t\left[v_t v_t'\right] < \infty$ (by Assumption C),

$$cov_{t,s}\left(v_t, v_s'\right) = 0,\ \ s \neq t.$$

This also implies that in the signal-noise decomposition, (5), v_t is white noise. Using this fact, Assumption B, Assumption C, and the decomposition of ζ_t above, it is immediate that, for $s \neq t$,

$$\begin{aligned} E_{t,s}\left[\zeta_t \zeta_s'\right] = {} & L\left(H \otimes H\right) G E_{t,s}\left[\sigma_t^2 \sigma_s^{2'}\right] G'\left(H \otimes H\right)' L' \\ & + L\left(H \otimes H\right) G E_{t,s}\left[\sigma_t^2 v_s'\right] + E_{t,s}\left[v_t \sigma_s^{2'}\right] G'\left(H \otimes H\right)' L'. \end{aligned} \tag{15}$$

By the law of iterated expectations, Assumption A.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 (15) vanishes, leaving

$$E_{t,s}\left[\zeta_t \zeta_s'\right] = L\left(H \otimes H\right) G E_{t,s}\left[\sigma_t^2 \sigma_s^{2'}\right] G'\left(H \otimes H\right)' L' + L\left(H \otimes H\right) G E_{t,s}\left[\sigma_t^2 v_s'\right]. \tag{16}$$

Finally, I can rewrite (16) as

$$\begin{aligned} & L\left(H \otimes H\right)\left(G E_{t,s}\left[\sigma_t^2 \sigma_s^{2'}\right] G' + G E_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s \varepsilon_s' - \Sigma_s\right)'\right]\right)\left(H \otimes H\right)' L' \\ & = L\left(H \otimes H\right) G \tilde{M}_{t,s}\left(H \otimes H\right)' L' \end{aligned} \tag{17}$$

where $\tilde{M}_{t,s} = E_{t,s}\left[\sigma_t^2 \sigma_s^{2\prime}\right] G' + E_{t,s}\left[\sigma_t^2 vec\left(\varepsilon_s \varepsilon_s' - \Sigma_s\right)'\right]$. $\tilde{M}_{t,s}$ is an $n \times n^2$ matrix. Since the same arguments can be applied to show that

$$E_t\left[\zeta_t\right] E_s\left[\zeta_t\right]' = L\left(H \otimes H\right)\left(G E_t\left[\sigma_t^2\right] E_s\left[\sigma_s^2\right] G'\right)\left(H \otimes H\right)' L',$$

it is immediate that additionally

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

where

$$M_{t,s} = \tilde{M}_{t,s} - E_t\left[\sigma_t^2\right] E_s\left[\sigma_s^{2\prime}\right] G'.$$

□

A.2 Proof of Theorem 1

I begin by proving two lemmata for properties of the singular value decomposition (SVD).[34]

Definition 1. Define

1. $U_1 D_U U_2' = V$, a reduced SVD, with V $n_1 \times n_2$, D_U $d \times d$,[35]
2. C_i is a full rank matrix, $m_i \times n_i, m_i \geq n_i$,
3. $F = C_1 V C_2'$, with $rank\left(F\right) = d$.

First, I show that a linear relationship exists between the singular vectors of V (U_1, which will later correspond to an unobservable object) and singular vectors of F (which will later correspond to an observable object).

Lemma 1. *There exists a matrix Γ_1 such that $C_1 U_1 \Gamma_1$ is an orthogonal matrix of singular vectors from an SVD of F.*

Proof. Define $Q_1 R_1 = C_1 U_1$, a QR decomposition, and similarly for $U_2 C_2$. Then F can be factored as $F = Q_1 R_1 D_U R_2' Q_2'$. The upper-triangular matrix R_1 is $d \times d$ and full rank since $C_1 U_1$ is full rank d ($rank\left(R_1\right) \geq rank\left(Q_1 R_1\right) = rank\left(C_1 U_1\right) \geq rank\left(F\right) = d$). Now take

[34]For a real-valued matrix V, the singular value decomposition $V = U_1 D_U U_2$ decomposes V into two orthogonal matrices U_1, U_2, and a non-negative diagonal matrix D_U. The "singular vectors", columns of U_1 and U_2, are eigenvectors of VV' and $V'V$ respectively. The non-zero singular values (diagonal of D_U) are square-roots of the non-zero eigenvalues of VV' and $V'V$.

[35]A reduced SVD reduces the dimension of U_1, D_U, U_2 to drop singular values equal to zero and their corresponding arbitrary singular vectors.

another singular value decomposition, this time of $R_1 D_U R_2'$, so $W_1 D_R W_2' = R_1 D_U R_2'$. Then F can be factored as $F = (Q_1 W_1) D_R (W_2' Q_2')$, which is itself a reduced SVD, since it can easily be shown that D_R are singular values of F, and the corresponding vectors are clearly orthogonal. To obtain Γ_1, recall $Q_1 R_1 = C_1 U_1$ and note $Q_1 R_1 \left(R_1^{-1} W_1\right) = Q_1 W_1$, singular vectors of F, so $\Gamma_1 = R_1^{-1} W_1$, which is guaranteed to exist. $\square$

Definition 2. Define $S_1 D_S S_2' = F$, a reduced SVD.

I now establish the uniqueness of an SVD of F up to orthogonal rotations, accounting for the possibility of repeated eigenvalues.

Lemma 2. *The SVD of F is unique up to rotations characterized by $F = S_1 B_1 D_S B_2 S_2'$ where B_i is orthogonal.*

Proof. For non-repeated singular values in D_S, the corresponding singular vectors are unique up to sign, and the space of vectors corresponding to any k repeated singular values corresponds to linear combinations of any k such vectors. Thus any alternative reduced singular value decomposition of F can be written as $F = (S_1 B_1) D_S (B_2 S_2')$, since B_i can incorporate any such sign changes or linear combinations. Since $S_i B_i$ must be orthogonal (by definition of an SVD), $B_i' S_i' S_i B_i = I_d$. Then since S_i is orthogonal, $B_i' B_i = I_d$, so B_i is itself orthogonal. $\square$

Definition 3. Define

1. $C_1 = (H \otimes H) G$, $n^2 \times n$ with rank n, $C_2 = (H \otimes H)$, $n^2 \times n^2$ with rank n^2,
2. G is a selection matrix such that $vec(ADA') = (A \otimes A) Gdiag(D)$,
3. $\tilde{S}_1 = C_1 \tilde{U}_1 = C_1 U_1 \Gamma_1 B_1$, singular vectors from any reduced SVD of F,
4. V is $n \times n^2$ and has no proportional rows,
5. $rank(V) = d \geq 2$.

Using the relationships I have derived in Lemma 1 between an SVD of the observable F and an SVD of the unobservable V, I now establish conditions under which H is uniquely determined from singular vectors of F. Using Lemma 2, I show that this is true even in the case of repeated singular values.

Proposition 3. *H is uniquely determined up to column order from the equations $F = C_1 V C_2'$ provided V has no proportional rows.*

Proof. U_1 is $n \times d$. Note $C_1 U_1 = \left[vec\left(Hdiag\left(U_1^{(1)}\right)H'\right), \ldots, vec\left(Hdiag\left(U_1^{(d)}\right)H'\right)\right]$, where $d \geq 2$ (this follows from the structure of C_1). By the proportional row condition on V, for any j, 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 $i = 1, 2, ..., n,\ i \neq j$ (since if V has no proportional rows, its left singular vectors also have no proportional rows). By an argument due to **?** (the underlying mathematical result also features in **?** and **?**), $H^{(j)}$ 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, provided $U_{1,j}^{(l)}/U_{1,i}^{(l)} \neq U_{1,j}^{(k)}/U_{1,i}^{(k)}$.

The same argument applies to $C_1\tilde{U}_1$ where $\tilde{U}_1 = U_1\Gamma_1 B_1$, provided $\tilde{U}_1$ has no proportional rows. To establish this, take any two rows in U_1 that are not proportional rows; multiplying by full rank Γ_1 cannot decrease their rank (so they cannot become proportional). The same holds true for multiplication by the orthogonal B_1. Thus H remains the unique solution to $C_1\tilde{U}_1$ up to column order (ordering of eigenvalues); for any permutation of H (ordering of eigenvalues), the scale of each column (eigenvector) is pinned down by the unit diagonal normalization of H. In other words, H is unique up to permutation and the associated renormalization (rescaling eigenvalues to maintain the unit diagonal of H). □

Proposition 3 is rewritten in terms of the identifying equations, setting $V = M_{t,s}$ and $F = cov_{t,s}(\zeta_t, \zeta_s')$ (ignoring L, which simply deletes duplicated equations) to yield Theorem 1, noting that the requirements imposed on U_1 imply the stated conditions on $M_{t,s}$. H is identified up to column permutations and the associated renormalization to maintain a unit diagonal. In other words, H is identified up to transformations given by $HP_i \times id(HP_i)$, where P_i is a permutation matrix.

A.3 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 $i = 1, 2, ..., n$. □

A.4 Proof of Theorem 2

Proof. Theorem 2 is based on the argument underlying Proposition 3. Note that $E_t[\zeta_t] = Lvec\left(Hdiag\left(E_t\left[\sigma_t^2\right]\right)H'\right) = L(H \otimes H)GE_t\left[\sigma_t^2\right] = LC_1E_t\left[\sigma_t^2\right]$. Recall that the equations used to identify H in Proposition 3, $S_1 = C_1U_1$, have identical form ($vec\left(Hdiag\left(U_1^{(i)}\right)H'\right), i = 1, 2, \ldots, n$). Thus (ignoring L, which simply deletes duplicated equations), the moment $E_t[\zeta_t]$ augments one additional column of identifying equations, so H may now be identified from $C_1\left[\begin{array}{cc} U_1 & E_t\left[\sigma_t^2\right] \end{array}\right] \equiv C_1\bar{U}$. Now, for identification to hold, there must be at least one k, l

pair for j such that $\bar{U}_j^{(l)}/\bar{U}_i^{(l)} \neq \bar{U}_j^{(k)}/\bar{U}_i^{(k)}$ for all $i = 1, 2, ..., n+1$ $i \neq j$. The "no proportional rows" condition that applied to U_1 in Theorem 1 now applies to $\bar{U}$; this condition will hold if it is also satisfied by $\bar{M}_{t,s} \equiv \begin{bmatrix} M_{t,s} & E_t\left[\sigma_t^2\right] \end{bmatrix}$. Note that the same logic can be extended to adding additional autocovariances, etc., in each case adding columns to $\bar{U}$, and thus decreasing the plausibility of the condition failing. $\square$

A.5 Proof of Corollary 2

Proof. Corollary 2 follows directly from Theorem 2 by noting that the $rank\left(\bar{M}_{t,s}\right) \geq 2$ is satisfied given even one dimension of time-varying volatility and re-arranging the condition under which row proportionality fails. $\square$

A.6 Proof of Proposition 2

Remark. The relationship between $\tilde{M}_{t,t-p}$, $M_{t,t-p}$, and $\bar{M}_{t.t-p} = \begin{bmatrix} M_{t,t-p} & E\left[\sigma_t^2\right] \end{bmatrix}$, which suggests the use of $\tilde{M}_{t,t-p}$ for the identification test, warrants further discussion. Given the broadest identification result, Theorem 2, applies rank conditions to $\bar{M}_{t.t-p}$, that is the ideal target for a test. However, given the structure of the problem, there is no reduced form object that inherits the rank of $\bar{M}_{t.t-p}$ directly. However, note that $rank\left(\bar{M}_{t.t-p}\right) = rank\left(\begin{bmatrix} \tilde{M}_{t,t-p} & E\left[\sigma_t^2\right] \end{bmatrix}\right)$, since $M_{t,t-p}$ is a linear combination of $\tilde{M}_{t,t-p}$ and $E\left[\sigma_t^2\right]$. It follows that $rank\left(\tilde{M}_{t,t-p}\right) \leq rank\left(\bar{M}_{t.t-p}\right)$ and similarly $rank\left(M_{t,t-p}\right) \leq rank\left(\bar{M}_{t.t-p}\right)$. This raises the question of whether $\tilde{M}_{t,t-p}$ or $M_{t,t-p}$ provides the better indication of the rank of $\bar{M}_{t.t-p}$. The rank of each will be at most 1 less than that of $\bar{M}_{t.t-p}$. However, $M_{t,t-p}$ will have lower rank whenever there is a homoskedastic series, a series with non-persistent stochastic variance, or two series whose variances constitute a linear transformation. Conversely, in each of these cases, the rank of $\tilde{M}_{t,t-p}$ will be equal to that of $\bar{M}_{t.t-p}$.[36] Thus, focusing on $\tilde{M}_{t,t-p}$ for rank tests allows for a more powerful test of identification in these key settings.

Proof. I begin by showing $rank\left(E\left[\zeta_t\zeta_{t-p}'\right]\right) = r$ if and only if $rank\left(\tilde{M}_{t,t-p}\right) = r$. Recall $E\left[\zeta_t\zeta_{t-p}'\right] = L\left(H \otimes H\right) G\tilde{M}_{t,t-p}\left(H \otimes H\right)' L'$ (see A.1). The elimination matrix L merely deletes repeated rows (and L' columns), so cannot impact rank. Thus it suffices to work with $\left(H \otimes H\right) G\tilde{M}_{t,t-p}\left(H \otimes H\right)'$. Denote $C = \left(H \otimes H\right)$, which is square with full rank n^2, since H is full rank n. G is a full rank $n^2 \times n$ matrix. First, if $rank\left(\tilde{M}_{t,t-p}\right) = r$, $rank\left(G\tilde{M}_{t,t-p}\right) = rank\left(\tilde{M}_{t,t-p}\right) = r$ since G is rank n. Because C is full rank and square, $rank\left(CG\tilde{M}_{t,t-p}\right) = rank\left(G\tilde{M}_{t,t-p}\right) = r$, and likewise $rank\left(CG\tilde{M}_{t,t-p}C'\right) = r$.

[36] While cases with the opposite ordering exist, they do so only for nonlinear models.

Thus, $rank\left(\tilde{M}_{t,t-p}\right)=r$ implies $rank\left(LCG\tilde{M}_{t,t-p}C'L'\right)=r$. Going the other way, if $rank\left(CG\tilde{M}_{t,t-p}C'\right)=r$, then $rank\left(CG\tilde{M}_{t,t-p}\right)=r$ since C' is full rank and square. For the same reason, it then follows that $rank\left(G\tilde{M}_{t,t-p}\right)=r$. Because G has rank n, it further follows that $rank\left(\tilde{M}_{t,t-p}\right)=r$. Thus, $rank\left(LCG\tilde{M}_{t,t-p}C'L'\right)=r$ implies $rank\left(\tilde{M}_{t,t-p}\right)=r$, so $rank\left(E\left[\zeta_t\zeta'_{t-p}\right]\right)=r$ if and only if $rank\left(\tilde{M}_{t,t-p}\right)=r$.

This means that if $rank\left(E\left[\zeta_t\zeta'_{t-p}\right]\right)=r=n$, then $rank\left(\tilde{M}_{t,t-p}\right)=n$. In that case, $rank\left(\tilde{M}_{t,t-p}\right)\geq 2$, satisfying the first identification condition in Theorem 2. Moreover, since $\tilde{M}_{t,t-p}$ is $n\times n^2$, it is full rank, so it must have no proportional rows, satisfying the second identification condition. $\bar{M}_{t.t-p}$ must also satisfy the conditions of the theorem, since its rank is weakly greater than that of $\tilde{M}_{t,t-p}$. □

A.7 Proof of Theorem 3

Proof. This is a direct restatement of a main result of **?**. If $E\widehat{\left[\zeta_t\zeta'_{t-p}\right]}$ is a consistent and asymptotically normal estimator of $E\left[\zeta_t\zeta'_{t-p}\right]$, $rank\left(E\left[\zeta_t\zeta'_{t-p}\right]\right)<\left(n^2+n\right)/2$ (which it is because the maximum rank of $\tilde{M}_{t,t-p}$ is n) and the finite sample estimate $E\widehat{\left[\zeta_t\zeta'_{t-p}\right]}$ is almost surely full rank, then their Assumption 1 is satisfied. Then the Cragg-Donald statistic and its limiting distribution are given in equation (9) of that paper. □

A.8 Proof of Theorem 4

Proof. Since $\hat{H}\xrightarrow{p}H_P$, $\hat{H}P_i\times id\left(\hat{H}P_i\right)\xrightarrow{p}H_PP_i\times id\left(H_PP_i\right)$ by the continuous mapping theorem (since P_i is just a matrix of zeros and ones and $id\left(\cdot\right)$ simply inverts scalar entries of $\hat{H}$). If $f\left(\cdot\right)$ is continuous, then $f\left(\hat{H}P_i\times id\left(\hat{H}P_i\right)\right)\xrightarrow{p}f\left(H_PP_i\times id\left(H_PP_i\right)\right)$, again by the continuous mapping theorem. Since the norm ρ is also a continuous function of $f\left(\cdot\right)$, similarly $\forall P_i,\ \mathcal{L}_f\left(\hat{H},P_i,f_0\right)\xrightarrow{p}\mathcal{L}_f\left(H_P,P_i,f_0\right)$, proving the first part of the theorem. $\hat{P}$ is defined as the permutation matrix minimizing $\mathcal{L}_f\left(\hat{H},P_i,f_0\right)$. Since by definition $\forall P_i\neq I_n,\ \mathcal{L}_f\left(H,I_n,f_0\right)<\mathcal{L}_f\left(H,P_i,f_0\right)$, there exists $\epsilon>0$ such that

$$\forall P_i\neq I_n,\ \epsilon<\mathcal{L}_f\left(H,P_i,f_0\right)-\mathcal{L}_f\left(H,I_n,f_0\right).$$

Since $\forall P_i,\ \mathcal{L}_f\left(\hat{H},P_i,f_0\right)\xrightarrow{p}\mathcal{L}_f\left(H_P,P_i,f_0\right)$, for any $\delta>0$ there exists T^* such that for $T>T^*$,

$$\sup_{P_i\in\mathcal{P}}\Pr\left(\left|\mathcal{L}_f\left(\hat{H},P_i,f_0\right)-\mathcal{L}_f\left(H_P,P_i,f_0\right)\right|>\epsilon/2\right)<\delta,$$

and thus, since $\mathcal{L}_f(H_P, P^*, f_0) = \mathcal{L}_f(H, I_n, f_0)$

$$\forall P_i \neq P^*,\ \Pr\left(\mathcal{L}_f\left(\hat{H}, P_i, f_0\right) \leq \mathcal{L}_f\left(\hat{H}, P^*, f_0\right)\right) < \delta$$

so

$$\forall P_i \neq P^*,\ \Pr\left(\mathcal{L}_f\left(\hat{H}, P_i, f_0\right) > \mathcal{L}_f\left(\hat{H}, P^*, f_0\right)\right) \to 1. \tag{18}$$

Since $\hat{P}$ is defined as the minimizer of $\mathcal{L}_f\left(\hat{H}, P_i, f_0\right)$, (18) implies that $\hat{P} \xrightarrow{p} P^*$. This establishes the second point of the theorem. Applying the continuous mapping theorem again, $\hat{H}_{lab} = \hat{H}\hat{P} \times id\left(\hat{H}\hat{P}\right) \xrightarrow{p} H_P P^* \times id\left(H_P P^*\right) = H$, confirming the third point of the theorem.

It remains to show that asymptotic inference on $\hat{H}_{lab}$ is unaffected by the labeling step, in which $\hat{P}$ is estimated. Let $\mathcal{G}_T(x)$ denote the distribution of $\hat{H}$ around H_P, so that

$$\Pr\left(T^{1/2}\left(\hat{H} - H_P\right) \leq x\right) = \mathcal{G}_T(x),$$

and, by the asymptotic normality assumption on $\hat{H}$, $\mathcal{G}(x) = \lim_{T\to\infty} \mathcal{G}_T(x)$ exists. Additionally, define

$$\mathcal{G}_{iT}(x) = \Pr\left(T^{1/2}\left(\left(\hat{H} - H_P\right)P_i \times id\left(\left(\hat{H} - H_P\right)P_i\right)\right) \leq x\right),$$

which is the distribution of a continuous function of $\left(\hat{H} - H_P\right)$. Then the distribution of $\hat{H}_{lab}$ can be written as

$$\Pr\left(T^{1/2}\left(\hat{H}_{lab} - H\right) \leq x\right) = \sum_{P_i \in \mathcal{P}} \Pr\left(P_i = \hat{P}\right)\mathcal{G}_T(x). \tag{19}$$

To obtain the asymptotic distribution, taking the limit of (19) yields

$$\begin{aligned}
\lim_{T\to\infty} \Pr\left(T^{1/2}\left(\hat{H}_{lab} - H\right) \leq x\right) &= \sum_{P_i \in \mathcal{P}} \lim_{T\to\infty}\left(\Pr\left(P_i = \hat{P}\right)\mathcal{G}_{iT}(x)\right) \\
&= \sum_{P_i \in \mathcal{P}} \mathbf{1}\left[P_i = P^*\right] \lim_{T\to\infty} \mathcal{G}_{iT}(x) \\
&= \lim_{T\to\infty} \mathcal{G}_{*T}(x),
\end{aligned}$$

where the second equality uses the fact that the limits of both terms on the right hand side exist and that $\hat{P}$ is consistent for P^*, and $\mathcal{G}_{*T}(\cdot)$ is the distribution associated with the permutation P^*. Since $\mathcal{G}(x)$, the limiting distribution of $\left(\hat{H} - H_P\right)$, is Gaussian (by assumption), and $\mathcal{G}_{iT}(x)$ is the distribution of a continuous function of $\left(\hat{H} - H_P\right)$, then $\mathcal{G}_*(x) = \lim_{T\to\infty} \mathcal{G}_{*T}(x)$ can be obtained directly using the delta method. Note that the function to which the delta method is applied consists simply of permutation and renormalization of the elements of $\hat{H}$, so the labeling problem does not affect the asymptotic distribution of $\hat{H}_{lab}$, beyond accounting for permutation and renormalization. $\square$

While I do not pursue it here, it is straightforward to extend these results to discontinuous $f(\cdot)$ (like several discussed in Section 5 of the Supplement) by establishing the consistency of $f(\cdot)$ without using the continuous mapping theorem.