Instrumental Variable Identification of Dynamic Variance Decompositions*

Mikkel Plagborg-Møller
Princeton University

Christian K. Wolf
Princeton University

August 10, 2017

Abstract: Empirical macroeconomists often estimate impulse response functions using external instruments (proxy variables) for the shocks of interest. However, existing methods do not answer the key question of how important the shocks are in driving macro aggregates, unless the researcher is willing to assume either a Structural Vector Autoregressive representation or that the shocks are directly observed. We provide tools for doing inference on forecast variance decompositions in a general semiparametric moving average model, disciplined only through the availability of valid external instruments. We show that the share of the forecast variance that can be attributed to a shock is partially identified, albeit with informative bounds. Point identification can be achieved under a shock recoverability assumption that is restrictive but weaker than invertibility (i.e., the SVAR model assumption). The degree of invertibility is also set-identified; hence, the invertibility assumption is testable. To perform inference, we construct easily computable, partial identification robust confidence intervals. Finally, we interpret our results through the lens of a workhorse structural macro model.

Keywords: external instrument, impulse response function, invertibility, proxy variable, variance decomposition. *JEL codes:* C32, C36.

---

*Email: mikkelpm@princeton.edu and ckwolf@princeton.edu. We received helpful comments from Isaiah Andrews, Pepe Montiel Olea, Jim Stock, Harald Uhlig, and Mark Watson. The first draft of this paper was written while Wolf was visiting the Bundesbank, whose hospitality is gratefully acknowledged.

1 Introduction

Empirical macroeconomists increasingly seek to estimate impulse response functions without relying on dubious functional form assumptions or identifying restrictions. For example, local projections (LP) have become a popular direct regression-based alternative to Vector Autoregression (VAR) methods (Jordà, 2005; Angrist et al., 2017). Additionally, instrumental variable (IV, also known as proxy variable) methods are now routinely used to conduct structural analysis under plausible identifying assumptions (Stock, 2008; Stock & Watson, 2012; Mertens & Ravn, 2013; Gertler & Karadi, 2015; Stock & Watson, 2016; Caldara & Kamps, 2017). The combination of these two ideas leads to an appealingly semiparametric method, LP-IV, with an economically transparent framework for identification (Mertens, 2015; Ramey, 2016; Barnichon & Brownlees, 2017; Stock & Watson, 2017).

While existing robust methods consistently estimate impulse response functions, there currently exists no general way to quantify how *important* a shock is in driving the fluctuations of observed macro time series. In the traditional structural VAR (SVAR) literature, this question was often central to the analysis and was routinely addressed using forecast variance decompositions. A forecast variance decomposition measures the fraction of the overall forecast variance for a variable that can be attributed to each of the driving shocks. Although crucial to understanding the causes of economic fluctuations, tools for doing inference on variance decompositions have hitherto been unavailable in the semiparametric LP-IV setting. Applied researchers have thus faced an unfortunate dilemma between a need to quantify the importance of shocks and the desire to avoid imposing a restrictive SVAR structure (or assuming that shocks are directly observed).

In this paper, we show precisely to what extent the data are informative about the importance of shocks in a general linear dynamic model with IVs. Our model allows for a general semiparametric moving average structure of shock transmission, consistent with essentially all linearized structural macroeconomic models. Assuming only validity of the instruments, we derive sharp – and informative – bounds on various definitions of the forecast variance decomposition. Point identification can be achieved under a further assumption that the shock of interest is recoverable from the infinite past, present, and future of the endogenous macro variables, a weaker condition than the often questionable invertibility requirement of SVAR analysis. We further sharply characterize the extent to which the data are informative about the degree of noninvertibility of the shock of interest. We illustrate the identification results through the lens of the popular Smets & Wouters (2007) model.

Finally, to perform inference, we develop easily computable, partial identification robust confidence intervals for forecast variance decompositions and other objects of interest.

Following Stock & Watson (2017), the LP-IV model that we analyze, although linear, is semiparametric in the sense that we allow for a completely general infinite moving average structure for the transmission of shocks to observed variables.[1] Our sole assumption on the IVs is the usual exclusion restriction – the IVs correlate with the shock of interest, but not the other shocks. Importantly, we allow the number of underlying exogenous shocks to be unknown and potentially exceed the number of observed endogenous variables. Unlike standard SVAR models, we do not restrict the shocks to be *invertible*, i.e., spanned by past and current (but not future) values of the observed endogenous variables.

In this baseline LP-IV model, we show that forecast variance decompositions are only partially identified, albeit with informative bounds. Hence, even with an infinite sample, it would be impossible to pinpoint the exact importance of the shock of interest. The identified set is an interval, with nontrivial lower and upper bounds computable from the joint spectral density of the macro variables and the IV. The bounds depend on the strength of the external IV and the informativeness of the observed macro variables about the shock of interest. We consider two definitions of the forecast variance decomposition that differ in the information set used for the forecast: either conditional on all past endogenous variables, or conditional on all past shocks. The identified sets for these two concepts differ, unless invertibility is imposed *a priori*.

As the LP-IV model does not assume that shocks are invertible *a priori*, we are able to sharply characterize the extent to which the data are informative about the degree of noninvertibility. Inference about invertibility is useful for gauging the ability of VAR models to perform valid structural analysis and for distinguishing between different classes of structural models, such as models with anticipated versus surprise shocks. The degree of *non*invertibility of a shock is inversely related to the R^2 in an (infeasible) regression of the shock on past and current values of the endogenous variables. We show that this R^2 measure is partially identified, and we discuss which properties of the data would allow us to either confirm or reject invertibility with certainty. Our main finding is that the data rule out invertibility if and only if the IV Granger-causes the observed endogenous variables.

Although the baseline model is partially identified, we additionally provide assumptions that guarantee point identification of certain variance decompositions and the degree of non-

[1] This Structural Vector Moving Average Model has been analyzed recently from a Bayesian viewpoint by Barnichon & Matthes (2017) and Plagborg-Møller (2017), although with little emphasis on IVs.

invertibility. Point identification obtains if the shock of interest is *recoverable*, i.e., spanned by the infinite past, present, and future of the endogenous macro variables. This assumption also yields point identification of historical decompositions, which account for the part of the realized data that can be attributed to the shock of interest. The recoverability condition – although restrictive – is satisfied in certain classes of macro models, such as news and noise shock models, and it is substantially weaker than the invertibility condition that is automatically, if unintentionally, assumed in SVAR analysis. In particular, it is automatically satisfied if there are as many variables as shocks – a necessary, but not sufficient condition for the usual invertibility requirement. Alternatively, point identification obtains if the IV is assumed to be perfect, i.e., proportional to the true shock, or if the exogenous noise parts of the different IVs are assumed to be uncorrelated. Still, we stress that researchers do not need to adopt any of these auxiliary assumptions to *partially* identify variance decompositions.

To make our identification analysis practically useful, we develop partial identification robust confidence intervals for all objects of interest. In a first step, the researcher estimates a reduced-form VAR jointly in the macro variables and IVs. To be clear, this step merely uses the reduced-form VAR as a convenient tool for approximating the second moments of the data; it does not assume an underlying *structural* VAR model with invertible shocks. The second step then constructs sample analogues of our population partial identification bounds and inserts these into the confidence procedure of Imbens & Manski (2004) and Stoye (2009). We then construct confidence intervals both for the unknown parameters and for the identified sets, including a Bayesian implementation of the latter. Our confidence intervals have asymptotically valid frequentist coverage under weak conditions. We also discuss a test of invertibility that has power against all noninvertible alternatives.

We illustrate the usefulness of our identification bounds through the lens of the well-known structural business cycle model of Smets & Wouters (2007). We assume that the econometrician observes aggregate output, inflation, and a short-term policy interest rate, but she does not exploit the underlying structure of the model for inference. We separately consider external instruments for three different shocks: a standard monetary policy shock, a forward guidance (anticipated monetary) shock, and a technology shock. These three shocks vary greatly in terms of their degree of invertibility and recoverability, and we show that invertibility-based (e.g., SVAR) identification of the latter two shocks is severely biased. Nevertheless, our partial identification bounds are informative in all cases, provided the IV is not weak. This result is particularly striking for the technology shock, since the macro aggregates provide little information about the short- or medium-run cycles of this shock.

Literature. Applied macroeconomists now routinely estimate impulse response functions by direct regressions (local projections), although this requires the strong assumption that the economic shock is directly observed (or can be estimated consistently). Ramey (2016) provides a survey. External IV (also known as proxy variable) methods are designed to avoid the often implausible assumption of directly observable shocks. While macroeconometric IV methods were originally developed for VARs (Stock, 2008; Stock & Watson, 2012; Mertens & Ravn, 2013), they have recently been imported into the LP framework, resulting in LP-IV.

A rapidly growing literature has provided inference tools for the LP-IV model, although variance decompositions have been neglected. The theoretical background for LP-IV estimation of impulse response functions was established by Mertens (2015), Ramey (2016), Barnichon & Brownlees (2017), Jordà et al. (2017), Ramey & Zubairy (2017), and most comprehensively Stock & Watson (2017). As we emphasize below, the usual IV assumption only allows for point identification of *relative* impulse responses, e.g., the responses of the macro variables to a shock which raises the first variable by 1 unit. We go further and derive the identified set of all LP-IV model parameters, including those that enter into variance decompositions. Variance decompositions are frequently reported in SVAR analysis, where identification is straight-forward due to the implicit invertibility assumption (Kilian & Lütkepohl, 2017, Ch. 4). Stock & Watson (2017) assume invertibility of all shocks to identify forecast variance decompositions and historical decompositions in an LP-IV model; we substantially strengthen this result by showing that recoverability of the shock of interest is sufficient to yield point identification of some of these objects.

The ability of the LP-IV framework to allow for noninvertible shocks is a key attraction. As is well known, the standard SVAR model imposes invertibility *a priori*. Because noninvertibility is a frequent feature of structural macro models, the issue has received a lot of attention in the SVAR literature (cf. references in Plagborg-Møller, 2017, Sec. 2.3). Stock & Watson (2017) develop an LP-IV-based test of noninvertibility. Our contribution in this area is to sharply characterize the identified set for the degree of noninvertibility of the shocks, which in turn shows under what conditions the data can conclusively reject either invertibility or noninvertibility.

Our sharp bounds for dynamic variance decompositions are reminiscent of but substantially different from Watson's (1993) method for quantifying the fit of a structural macro model. Specifically, our derivations rely on the same basic idea that the identified set must consist of parameters that yield a positive semidefinite joint spectrum for observed and unobserved stationary time series. However, Watson's model and question bear little relationship with ours, so his results cannot easily be translated to the LP-IV setting.

Our confidence interval methods combine the literatures on partial identification and VAR-based spectral estimation. We employ the general confidence interval construction for interval-identified parameters developed by Imbens & Manski (2004) and Stoye (2009). The idea of using a VAR to approximate the spectrum of observed variables is reminiscent of VAR-HAC procedures for estimating long-run variances (Den Haan & Levin, 1997). Our explicit focus on partial identification in the LP-IV model is inspired by and complementary to the recent literature on robust inference in sign-identified SVAR models (Moon et al., 2013; Giacomini & Kitagawa, 2015; Gafarov et al., 2017).

Outline. Section 2 defines the LP-IV model and the parameters of interest. Section 3 contains our main results on identification of variance decompositions and the degree of noninvertibility. We first derive the results in a static setting for illustration and then turn to the general dynamic model. Section 4 interprets the results through the lens of the well-known Smets & Wouters (2007) structural macro model. Section 5 develops partial identification robust confidence intervals. Appendix A contains supplementary results and formulas, while proofs and auxiliary lemmas are relegated to Appendix B.

2 Model and parameters of interest

We begin by defining the Local Projection Instrumental Variable (LP-IV) model, its parameters of interest, and the notation we use throughout the paper. The LP-IV model allows for an unrestricted linear shock transmission mechanism and does not assume shocks to be invertible, unlike standard SVAR analysis. We assume the availability of valid external IVs (proxy variables) – variables that correlate with the shock of interest, but not with the other shocks. In addition to impulse response functions, our main objects of interest are forecast variance decompositions and the degree of noninvertibility of the shock of interest.

Model. We start out by describing the LP-IV model's semiparametric assumptions on shock transmission and the instrument exclusion restrictions. For notational clarity (and without loss of generality), we assume throughout that all time series below have mean zero and are free of any deterministic dynamics.

First, we specify the weak assumptions on shock transmission to endogenous variables. The n_y-dimensional vector $y_t = (y_{1,t}, \dots, y_{n_y,t})'$ of observed macro variables is driven by an unobserved n_ε-dimensional vector $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n_\varepsilon,t})'$ of exogenous economic shocks,

$$y_t = \Theta(L)\varepsilon_t, \quad \Theta(L) \equiv \sum_{\ell=0}^{\infty} \Theta_\ell L^\ell, \tag{1}$$

where L is the lag operator. The (i, j) element $\Theta_{i,j,\ell}$ of the $n_y \times n_\varepsilon$ moving average coefficient matrix Θ_ℓ is the impulse response of variable i to shock j at horizon ℓ. The j-th column of Θ_ℓ is denoted by $\Theta_{\bullet,j,\ell}$ and the i-th row by $\Theta_{i,\bullet,\ell}$. To obtain a nonsingular stochastic process, we assume that the polynomial $x \mapsto \det(\Theta(x))$ has no roots on the unit circle. This condition requires $n_\varepsilon \geq n_y$, but – crucially – we do not assume that the number of shocks n_ε is known. The model is semiparametric in that we place no *a priori* restrictions on the coefficients of the infinite moving average, except to ensure a valid stochastic process. In particular, we do not impose the usual invertibility conditions that point-identify $\Theta(L)$ in reduced-form time series analysis. It is well known that the infinite-order Structural Vector Moving Average model (1) is consistent with discrete-time Dynamic Stochastic General Equilibrium (DSGE) models as well as stable SVAR models for y_t. However, the principal appeal of LP-IV analysis is that it does not require an underlying SVAR structure or a fully-specified DSGE model.

Second, we assume the availability of external IVs for the shock of interest.[2] We specify the shock of interest to be the first one, $\varepsilon_{1,t}$. The n_z-dimensional vector $z_t = (z_{1,t}, \dots, z_{n_z,t})'$ of IVs is assumed to correlate with the first shock but not the other shocks,

$$z_t = \sum_{\ell=1}^{\infty} (\Psi_\ell z_{t-\ell} + \Lambda_\ell y_{t-\ell}) + \alpha\lambda\varepsilon_{1,t} + \Sigma_v^{1/2} v_t, \tag{2}$$

where Ψ_ℓ is $n_z \times n_z$, Λ_ℓ is $n_z \times n_y$, λ is an n_z-dimensional vector normalized to unit length ($\|\lambda\| = 1$) and with its first nonzero element being positive, $\alpha \geq 0$ is a scalar, and Σ_v is a symmetric positive semidefinite $n_z \times n_z$ matrix. Throughout the paper, $\|\cdot\|$ refers to the Euclidean norm. The key restriction on the IVs is that the shock of interest $\varepsilon_{1,t}$ is the only contemporaneous shock to enter into the equation (2). The scale parameter α (along with the residual variance-covariance matrix Σ_v) measures the overall strength of the IVs, while the unit-length vector λ determines which IVs are stronger than others.

Finally, we assume that the structural shocks and IV disturbances are jointly i.i.d. standard Gaussian,

$$(\varepsilon_t', v_t')' \overset{i.i.d.}{\sim} N(0, I_{n_\varepsilon + n_z}), \tag{3}$$

[2]If instruments are not available, the model (1) is severely underidentified (Lippi & Reichlin, 1994).

where I_n denotes the n-dimensional identity matrix. The mutual independence of the shocks is the standard assumption in empirical macroeconomics. We adopt the Gaussianity assumption for notational convenience. As usual, versions of all our results can be phrased using the language of linear projections.[3] The sole meaningful restriction is that we only consider identification from the second-moment properties of the data, as is standard in the applied macro literature (and without loss of generality for Gaussian data). Also note that we have normalized the variances of all shocks to 1. Some authors prefer a different parametrization of the LP-IV model where the shocks have non-unit variances but instead some impulse responses $\Theta_{i,j,\ell}$ are normalized to 1 (Stock & Watson, 2016, Sec. 4.1.3). All our results can be translated one-for-one into this alternative parametrization by simple scaling.

Throughout the paper we tacitly assume that the $(n_y + n_z)$-dimensional data vector $(y_t', z_t')'$ is strictly stationary. This is achieved by assuming that the elements of Θ_ℓ, Ψ_ℓ, and Λ_ℓ are absolutely summable across ℓ, and the polynomial $x \mapsto \det(I_{n_z} - \sum_{\ell=1}^{\infty} \Psi_\ell x^\ell)$ has all its roots outside the unit circle.

We allow for lagged values of z_t and y_t on the right-hand side of (2) because this is precisely enough to ensure that the LP-IV model is untestable (using second moments). In other words, there always exists a model of the form (1)–(3) that can match any given autocovariance structure of $(y_t', z_t')'$, cf. Proposition 1 below. If the economic application allows the researcher to exclude certain lagged terms *a priori*, this provides testable restrictions that are straight-forward to impose in our VAR-based inference procedures in Section 5. Except for the lag terms, our model for y_t and z_t is essentially identical to the LP-IV model studied by Stock & Watson (2017).

Invertibility and recoverability. We now define invertibility, the degree of noninvertibility, and recoverability.

The shock $\varepsilon_{1,t}$ is said to be *invertible* if it is spanned by past and current (but not future) values of the endogenous variables y_t: $\varepsilon_{1,t} = E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau\leq t})$. Invertibility of all structural shocks is assumed automatically by SVAR models, but the condition may or may not hold in a given moving average model (1), depending on the impulse response parameters Θ_ℓ. A sufficient condition for invertibility of all shocks is that the polynomial $x \mapsto \det(\Theta(x))$ has all its roots outside the unit circle. In many structural macro models, at least some of the shocks cannot be recovered from only past and current observed macro variables, i.e.,

[3] Replace conditional expectations by linear projections and replace conditional variances by variances of projection residuals.

the moving average representation is noninvertible. For example, this is often the case in models with news (anticipated) shocks or noise (signal extraction) shocks. Furthermore, if the number of structural shocks n_ε strictly exceeds the number of endogenous variables n_y, it is impossible for all shocks to be invertible.

A continuous measure of the degree of *non*invertibility is the R^2 value in a population regression of the shock on past and current observed variables. More generally, define

$$R_\ell^2 \equiv \frac{\operatorname{Var}(\varepsilon_{1,t}) - \operatorname{Var}(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell})}{\operatorname{Var}(\varepsilon_{1,t})} = 1 - \operatorname{Var}(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell})$$

as an R^2 measure of noninvertibility of the shock of interest using data up to time $t+\ell$. If the shock is invertible in the sense of the previous paragraph, then $R_\ell^2 = 1$ for all $\ell \geq 0$. If $R_\ell^2 < 1$ for some $\ell \geq 0$, then the model is noninvertible and thus no SVAR model could generate the impulse responses $\Theta(L)$, although the model may be *nearly* consistent with an SVAR structure if the R^2 values are close to 1 (Sims & Zha, 2006, pp. 243–245; Wolf, 2017). For noninvertible models, a plot of R_ℓ^2 for $\ell = 0, 1, 2, \ldots$ reveals how quickly the econometrician learns about the structural shocks over time.

A weaker condition than invertibility is that the shock of interest is *recoverable* from all leads and lags of the endogenous variables – that is, if $E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau<\infty}) = \varepsilon_{1,t}$, or equivalently if $R_\infty^2 = 1$. This property will become important when we consider assumptions that guarantee point identification.

Variance decompositions/ratios. In addition to the impulse responses Θ_ℓ, a primary object of interest is the forecast variance decomposition of the moving average model. Forecast variance decompositions capture the share of the forecast variance that can be attributed to particular shocks. In other words, they provide a quantitative measure of the importance of different shocks in generating macroeconomic fluctuations.

We consider two variance decomposition concepts. First, define the *forecast variance ratio* (FVR) for the shock of interest for variable i at horizon ℓ as

$$FVR_{i,\ell} \equiv 1 - \frac{\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t}, \{\varepsilon_{1,\tau}\}_{t<\tau<\infty})}{\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})} = \frac{\sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2}{\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})}$$

The FVR measures the reduction in the forecast variance that would come from knowing the entire path of future realizations of the first shock. The larger this measure is, the more important is the first shock for forecasting variable i at horizon ℓ. The FVR is always between 0 and 1. An unappealing feature, however, is that the FVR conflates two different sources of forecasting uncertainty. Writing out the denominator:

$$\begin{aligned}\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t}) &= \operatorname{Var}\left(\sum_{m=0}^{\infty} \Theta_{i,\bullet,m}\varepsilon_{t+l-m} \;\middle|\; \{y_\tau\}_{-\infty<\tau\leq t}\right) \\ &= \sum_{m=0}^{\ell-1} \Theta_{i,\bullet,m}\Theta'_{i,\bullet,m} + \operatorname{Var}\left(\sum_{m=\ell}^{\infty} \Theta_{i,\bullet,m}\varepsilon_{t+l-m} \;\middle|\; \{y_\tau\}_{-\infty<\tau\leq t}\right).\end{aligned}$$

Thus, the denominator of the FVR captures both fundamental forecasting uncertainty (uncertainty related to future shock realizations) and noninvertibility-induced forecasting uncertainty (uncertainty related to imperfect knowledge about *past* shocks). In contrast, the numerator only reflects fundamental forecasting uncertainty. This means that, when the first shock is noninvertible, the FVR does not equal 1 even if the first shock is solely responsible for driving the i-th variable in equation (1).

The second variance decomposition concept is the *forecast variance decomposition* (FVD) for the shock of interest for variable i at horizon ℓ,

$$FVD_{i,\ell} \equiv 1 - \frac{\operatorname{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty<\tau\leq t}, \{\varepsilon_{1,\tau}\}_{t<\tau<\infty})}{\operatorname{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty<\tau\leq t})} = \frac{\sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2}{\sum_{j=1}^{n_\varepsilon}\sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2}. \tag{4}$$

The FVD measures the reduction in forecast variance that arises from learning the path of future realizations of the shock of interest, supposing that we already had the history of structural shocks ε_t available when forming our forecast. Because the econometrician generally does not observe the structural shocks directly, the FVD is best thought of as reflecting forecasts of economic agents who observe the underlying shocks. The FVD always lies between 0 and 1, purely reflects fundamental forecasting uncertainty, and equals 1 if the first shock is the only shock driving variable i in equation (1).

While the FVR and FVD concepts generally differ, they coincide in the case where all shocks are invertible, since in that case the information set $\{y_\tau\}_{-\infty<\tau\leq t}$ equals the information set $\{\varepsilon_\tau\}_{-\infty<\tau\leq t}$. This explains why the SVAR literature has not made the distinction between the two concepts.

3 Identification

This section presents our main results on instrumental variable identification of variance decompositions and the degree of noninvertibility. For exposition, we start by deriving results for a static version of the LP-IV model. We then turn to the general dynamic model, which applies the static results to the frequency domain representation of the data. The dynamics involve additional nuances in characterizing the informativeness of the macro aggregates for the shock at all frequencies. Our main results assume availability of a single external IV for the shock of interest. In the last subsection we show that identification analysis in a model with multiple IVs for the same shock can be reduced to the single-IV case without loss of generality, as long as the multiple-IV model is not rejected by the data.

3.1 Static model

We use an illustrative static model to motivate why variance decompositions are partially identified in the general case but can be point-identified under additional assumptions. Although the static model does not capture all the nuances of the dynamic LP-IV model, it provides useful intuition for the general case.

Model. The static model with a single IV assumes[4]

$$\begin{aligned} y_t &= \Theta_0 \varepsilon_t, \\ z_t &= \alpha \varepsilon_{1,t} + \sigma_v v_t, \\ (\varepsilon_t', v_t)' &\overset{i.i.d.}{\sim} N(0, I_{n_\varepsilon+1}), \end{aligned}$$

where Θ_0 is $n_y \times n_\varepsilon$, and $\alpha, \sigma_v \geq 0$ are scalars. To avoid singularity, we assume that Θ_0 has full row rank, so in particular $n_\varepsilon \geq n_y$.

In the static case, the degree of noninvertibility of the model is fully summarized by the static projection R^2, defined as

$$R_0^2 = 1 - \operatorname{Var}(\varepsilon_{1,t} \mid y_t) = \Theta_{\bullet,1,0}' \operatorname{Var}(y_t)^{-1} \Theta_{\bullet,1,0}.$$

Here the second equality follows from the usual linear projection formula.

As for variance decompositions, the static model does not distinguish between the FVR

[4] While the static model is primarily intended for gaining intuition, the results in this subsection are directly relevant for SVAR analysis with an external IV. In that framework, y_t would be the reduced-form VAR residuals, which are a linear function of the vector ε_t of contemporaneous structural shocks. Textbook SVAR analysis further assumes that $n_\varepsilon = n_y$, so the model is identified up to an orthogonal rotation matrix.

and FVD, and we can restrict attention to one-step prediction:

$$FVD_{i,1} = 1 - \frac{\operatorname{Var}(y_{i,t} \mid \varepsilon_{1,t})}{\operatorname{Var}(y_{i,t})} = \frac{\Theta_{i,1,0}^2}{\operatorname{Var}(y_{i,t})}.$$

Partial identification. We now show that the impulse response functions, the degree of noninvertibility, and variance decompositions are all identified up to a scalar multiple. This factor of proportionality is interval-identified, with nontrivial and informative lower and upper bounds. Since the data is i.i.d., identification in this model relies solely on contemporaneous covariance calculations.

It is immediate that impulse responses $\Theta_{i,1,0}$ to the shock of interest are identified up to the scale parameter α:

$$\operatorname{Cov}(y_t, z_t) = \alpha \Theta_{\bullet,1,0}. \tag{5}$$

In particular, relative responses are identified, cf. Stock & Watson (2017). Since the vector $\Theta_{\bullet,1,0}$ is identified up to scale α, the degree of noninvertibility R_0^2 is identified up to the multiple $\frac{1}{\alpha^2}$, and the FVDs of different variables i are identified up to the same multiple $\frac{1}{\alpha^2}$.

What values of the scale parameter α are consistent with the distribution of the data $w_t = (y_t', z_t)'$? First, the equation defining the IV z_t implies

$$\alpha^2 \leq \operatorname{Var}(z_t) \equiv \alpha_{UB}^2. \tag{6}$$

The boundary case $\alpha = \alpha_{UB}$ corresponds to $\operatorname{Var}(\varepsilon_{1,t} \mid z_t) = 0$, i.e., perfect informativeness of the instrument. Second, we find

$$\operatorname{Var}(E(z_t \mid y_t)) = \operatorname{Cov}(y_t, z_t)' \operatorname{Var}(y_t)^{-1} \operatorname{Cov}(y_t, z_t) = \alpha^2 e_1' \{\Theta_0' (\Theta_0 \Theta_0')^{-1} \Theta_0\} e_1, \tag{7}$$

where e_1 denotes the unit vector with 1 as the first element and zeros elsewhere. If we knew there were as many shocks as variables, $n_\varepsilon = n_y$, then Θ_0 would be square and invertible, and expression (7) would reduce to α^2.[5] However, more generally allowing for the possibility $n_\varepsilon > n_y$, the matrix in curly brackets in (7) is a projection matrix, whose eigenvalues are all 0 or 1. Hence, we always obtain a lower bound

$$\alpha^2 \geq \operatorname{Var}(E(z_t \mid y_t)) \equiv \alpha_{LB}^2. \tag{8}$$

[5]This calculation is frequently used in SVAR-IV analysis, cf. Mertens & Ravn (2013), Gertler & Karadi (2015), and Stock & Watson (2017).

The boundary case $\alpha = \alpha_{LB}$ corresponds to $\text{Var}(\varepsilon_{1,t} \mid y_t) = 0$, i.e., the observed macro aggregates y_t are perfectly informative about the hidden shock (invertibility).

It is not hard to show (and it follows from our general results below) that the bounds (6) and (8) on α^2 are *sharp*, in the following sense: Given any positive semidefinite variance-covariance matrix for $w_t = (y_t', z_t)'$, and given any value of α^2 in the interval between the bounds (6) and (8), we can construct a static model with the given value of α and which matches the given $\text{Var}(w_t)$ (we just have to choose Θ_0 and σ_v appropriately).

The width of the identified set $[\alpha_{LB}^2, \alpha_{UB}^2]$ for α^2 depends on the degree of noninvertibility and the strength of the instrument. The interval is never empty, and it collapses to a point only in the knife-edge case of a perfectly informative instrument *and* invertibility of the first shock. Generically, α – and so impulse responses, FVDs, and the degree of noninvertibility – are only partially identified, but with useful bounds that limit the range of admissible values. Intuitively, the bounds arise because a large α requires a large variance of the IV, while a small α requires y_t and z_t to be nearly uncorrelated.

To interpret the identified set of $\frac{1}{\alpha^2}$, we can express it in terms of the underlying (unknown) model parameters:

$$\left[\underbrace{\frac{\alpha^2}{\alpha^2 + \sigma_v^2}}_{\text{instrument strength}} \times \frac{1}{\alpha^2}, \quad \underbrace{\frac{1}{R_0^2}}_{\text{recoverability}} \times \frac{1}{\alpha^2} \right].$$

The lower bound is more informative (i.e., larger and closer to the true $\frac{1}{\alpha^2}$) when the instrument is stronger in the sense of a higher signal-to-noise ratio. Conversely, the upper bound is more informative (i.e., smaller and closer to the true $\frac{1}{\alpha^2}$) when the model is closer to being invertible for the shock of interest.

Having partially identified the scale parameter, we obtain identified sets for the FVD and degree of noninvertibility. By scaling the identified set for $\frac{1}{\alpha^2}$, we find the identified set for $FVD_{i,0}$:

$$\left[\underbrace{\frac{1}{\text{Var}(z_t)} \times \frac{\text{Cov}(y_{i,t}, z_t)^2}{\text{Var}(y_{i,t})}}_{\frac{\alpha^2}{\alpha^2+\sigma_v^2} \times FVD_{i,0}}, \quad \underbrace{\frac{1}{\text{Var}(E(z_t \mid y_t))} \times \frac{\text{Cov}(y_{i,t}, z_t)^2}{\text{Var}(y_{i,t})}}_{\frac{1}{R_0^2} \times FVD_{i,0}} \right].$$

Instrument informativeness and invertibility thus map one-to-one into the width of the identified set for the FVD. The identified set for the degree of noninvertibility R_0^2 can similarly be obtained by scaling the identified set for $\frac{1}{\alpha^2}$. The identified set for R_0^2 always contains 1, i.e., the data can never reject invertibility in the static model.

Sufficient conditions for point identification. Although the baseline model is partially identified, point identification obtains under a variety of auxiliary assumptions.

First, assume that the shock of interest is recoverable, which in the static model is the same as invertibility: $E(\varepsilon_{1,t} \mid y_t) = \varepsilon_{1,t}$, or equivalently $R_0^2 = 1$. Then α^2 is given by the lower bound in (8); we can then identify the impulse responses $\Theta_{\bullet,1,0}$ from the covariance relationship (5), and σ_v from $\text{Var}(z_t)$. Hence, all objects of interest are point-identified under the recoverability assumption. A stronger condition than recoverability is that there are as many shocks as variables, $n_\varepsilon = n_y$. This condition implies that Θ_0 is square and invertible, so *all* shocks are recoverable, and point identification follows.

Second, restrictions on the instrument z_t can also be sufficient to ensure point identification. Point identification obtains if the instrument is perfect; that is, if $\sigma_v = 0$. In that case the shock $\varepsilon_{1,t}$ is effectively observed by the econometrician and all parameters can be identified directly from regressions of $y_{i,t}$ on z_t (local projections). Equivalently, the true α is given by the upper bound in (6), and then all derivations follow as before. An alternative point-identifying assumption is the availability of multiple instruments with mutually independent "first-stage disturbances" v_t, as shown in Section 3.3.

3.2 General dynamic model

We now present our main identification results for the general dynamic model, applying the logic of the static model frequency-by-frequency to the frequency domain representation of the data. The main building block result is that, exactly as in the static model, the identified set for the scale parameter α is an interval with informative bounds. From this result we derive identified sets for the main objects of interest: the degree of noninvertibility and variance decompositions. Relative to the static case, the dynamic case involves additional nuances in characterizing the informativeness of the data for the hidden shock at all frequencies.

For the moment, we carry out the analysis for the case of a single IV ($n_z = 1$), leaving the generalization to Section 3.3. That is, z_t is a scalar and $\lambda = 1$ in equation (2). We write $\Sigma_v^{1/2} = \sigma_v \geq 0$, a scalar.

Preliminaries. For the identification analysis, it will prove convenient to define and work with the IV projection residual

$$\tilde{z}_t \equiv z_t - E(z_t \mid \{y_\tau, z_\tau\}_{-\infty<\tau<t}) = \alpha\varepsilon_{1,t} + \sigma_v v_t. \tag{9}$$

We have thus removed any dependence on lagged observed variables, and $\tilde{z}_t$ is serially uncorrelated by construction.

Next, we need to define our notation for spectral density matrices. For any two jointly stationary vector time series a_t and b_t of dimensions n_a and n_b, respectively, define the $n_a \times n_b$ cross-spectral density matrix function (Brockwell & Davis, 1991, Ch. 4 and 11)

$$s_{ab}(\omega) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} e^{-i\omega\ell} \operatorname{Cov}(a_t, b_{t-\ell}), \quad \omega \in [0, 2\pi].$$

This object is well-defined if the autocovariance function of $(a_t', b_t')'$ is absolutely summable. For any vector time series a_t, we denote its spectrum by $s_a(\omega) = s_{aa}(\omega)$.

Just like identification in the static case proceeded through the variance-covariance matrix of the data, identification in the general dynamic model will rely heavily on the joint spectrum for $w_t = (y_t', \tilde{z}_t)'$ implied by the LP-IV model, i.e., equations (1), (3), and (9). This joint spectrum is given by

$$s_w(\omega) = \begin{pmatrix} s_y(\omega) & s_{y\tilde{z}}(\omega) \\ s_{y\tilde{z}}(\omega)^* & s_{\tilde{z}}(\omega) \end{pmatrix} = \frac{1}{2\pi} \begin{pmatrix} \Theta(e^{-i\omega})\Theta(e^{-i\omega})^* & \alpha\Theta(e^{-i\omega})e_1 \\ \alpha e_1'\Theta(e^{i\omega})' & \alpha^2 + \sigma_v^2 \end{pmatrix}, \ \omega \in [0, 2\pi], \quad (10)$$

where an asterisk denotes complex conjugate transpose, e_1 is again the unit vector with 1 as the first element and zeros elsewhere, and the matrix polynomial $\Theta(\cdot)$ was defined in equation (1).[6] Note the similarity between the spectrum $s_w(\omega)$ and the covariance structure in the static model in Section 3.1. The main difference is that in the dynamic setting we have a matrix at each frequency $\omega \in [0, 2\pi]$.

Impulse responses. As in the static model, the impulse responses to the first shock are identified up to the scale parameter α:

$$\operatorname{Cov}(y_t, \tilde{z}_{t-\ell}) = \alpha\Theta_{\bullet,1,\ell}. \quad (11)$$

This is also clear from expression (10) for the cross-spectrum of y_t and $\tilde{z}_t$. Thus, *relative* impulse responses $\Theta_{i,1,\ell}/\Theta_{11,0}$ are identified, as shown by Stock & Watson (2017) and others.

Scale parameter. We now show that, exactly as in the static case, the identified set for α is an interval with informative bounds. Although α itself is not a parameter of primary

[6] See Brockwell & Davis (1991, Ch. 11) for references on spectral densities of moving average processes.

interest, the results in the next paragraphs are key ingredients to identification of variance decompositions and the degree of noninvertibility.

Just as in the static case, the variance of the instrument provides the upper bound:

$$\alpha^2 \leq \text{Var}(\tilde{z}_t) \equiv \alpha_{UB}^2. \tag{12}$$

As in the static model, the boundary case $\alpha = \alpha_{UB}$ corresponds to perfect instrument informativeness. To derive the lower bound, we apply the argument from the static case to the joint spectrum of the data at every frequency. For every $\omega \in [0, 2\pi]$,

$$2\pi s_{y\tilde{z}}(\omega)^* s_y(\omega)^{-1} s_{y\tilde{z}}(\omega) = \alpha^2 e_1' \left\{ \Theta(e^{-i\omega})^* \left(\Theta(e^{-i\omega}) \Theta(e^{-i\omega})^* \right)^{-1} \Theta(e^{-i\omega}) \right\} e_1 \leq \alpha^2. \tag{13}$$

The far left-hand side above equals $2\pi s_{\tilde{z}^\dagger}(\omega)$, where

$$\tilde{z}_t^\dagger \equiv E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau<\infty})$$

is the projection of $\tilde{z}_t$ onto all lags and leads of the endogenous variables y_t.[7] Hence, we obtain the lower bound[8]

$$\alpha^2 \geq 2\pi \sup\nolimits_{\omega \in [0,\pi]} s_{\tilde{z}^\dagger}(\omega) \equiv \alpha_{LB}^2. \tag{14}$$

This lower bound generalizes the lower bound (8) in the static model. Intuitively, in the static case a small value of α requires y_t and $\tilde{z}_t$ to be nearly independent. In the dynamic case, a small value of α requires $\tilde{z}_t$ to be nearly unpredictable by y_t at *every* frequency ω, e.g., both in the long run and at business cycle frequencies. The boundary case $\alpha = \alpha_{LB}$ corresponds to the observed macro aggregates being perfectly informative about the hidden shock $\varepsilon_{1,t}$ at *some* frequency $\overline{\omega} \in [0, \pi]$, i.e., $s_{\varepsilon_1^\dagger}(\overline{\omega}) \approx s_{\varepsilon_1}(\cdot) = \frac{1}{2\pi}$.

The main building block result of this paper is that the above bounds $\alpha_{LB}^2, \alpha_{UB}^2$ are *sharp*.

Proposition 1. *Let there be given a joint spectral density for $w_t = (y_t', \tilde{z}_t)'$, continuous and positive definite at every frequency, with $\tilde{z}_t$ being unpredictable from $\{w_\tau\}_{-\infty<\tau<t}$. Choose any $\alpha \in (\alpha_{LB}, \alpha_{UB}]$. Then there exists a model of the form* (1)*,* (3)*, and* (9) *with the given α such that the spectral density of w_t implied by the model matches the given spectral density.*

In words, the distribution of the data allows us to conclude that α^2 lies in the identified

---

[7] See Brockwell & Davis (1991, Remark 3, p. 439). They consider the bivariate case, but the multivariate generalization is straight-forward. Note that $s_{\tilde{z}^\dagger}(\omega)$ need not be constant in ω even though $s_{\tilde{z}}(\omega)$ is.

[8] As always with univariate spectral densities, $s_{\tilde{z}^\dagger}(2\pi - \omega) = s_{\tilde{z}^\dagger}(\omega)$ for $\omega \in [0, \pi]$.

set $[\alpha^2_{LB}, \alpha^2_{UB}]$, but the data cannot rule out any values of α^2 in this interval. The proposition does not cover the knife-edge case $\alpha = \alpha_{LB}$, which presents some economically inessential technical difficulties.

The width of the identified set for α^2 depends on the application, although the set is never empty. To interpret the identified set, we can express it in terms of the underlying (unknown) model parameters. To that end, define

$$\varepsilon^{\dagger}_{1,t} \equiv E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau<\infty}).$$

In almost perfect analogy to the static case, the identified set for $\frac{1}{\alpha^2}$ is then

$$\Bigg[\underbrace{\frac{\alpha^2}{\alpha^2+\sigma_v^2}}_{\text{instrument strength}} \times \frac{1}{\alpha^2}, \quad \underbrace{\frac{1}{2\pi \sup_{\omega\in[0,\pi]} s_{\varepsilon_1^\dagger}(\omega)}}_{\text{informativeness of data for shock}} \times \frac{1}{\alpha^2}\Bigg].$$

The lower bound of the identified set for $\frac{1}{\alpha^2}$ is larger (and closer to the true $\frac{1}{\alpha^2}$) when the instrument is stronger in the sense of a higher signal-to-noise ratio. The upper bound of the identified set for $\frac{1}{\alpha^2}$ is smaller (and closer to the true $\frac{1}{\alpha^2}$) when the data are more informative about the shock of interest at least at *some* frequency. Similar to the static case, the identified set for $\frac{1}{\alpha^2}$ does not collapse to a point unless the instrument is perfect *and* there exists a frequency $\overline{\omega}$ for which the data are perfectly informative about the frequency-$\overline{\omega}$ cyclical component of the shock.

To further interpret α^2_{LB}, we derive a lower bound to this object that is explicitly tied to the degree of non-recoverability/noninvertibility. First, we have

$$\alpha^2_{LB} = 2\pi \sup_{\omega\in[0,\pi]} s_{\tilde{z}^\dagger}(\omega) \geq \int_0^{2\pi} s_{\tilde{z}^\dagger}(\omega)\, d\omega = \operatorname{Var}(\tilde{z}^\dagger_t). \tag{15}$$

The far right-hand side above depends on the degree of non-recoverability of the shock:

$$\operatorname{Var}(\tilde{z}^\dagger_t) = \operatorname{Var}(E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau<\infty})) = \alpha^2(1 - \operatorname{Var}(\varepsilon_{1t} \mid \{y_\tau\}_{-\infty<\tau<\infty})) = \alpha^2 \times R^2_\infty.$$

An even lower bound on α^2_{LB} is given by

$$\operatorname{Var}(E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t})) = \alpha^2(1 - \operatorname{Var}(\varepsilon_{1t} \mid \{y_\tau\}_{-\infty<\tau\leq t})) = \alpha^2 \times R^2_0.$$

Thus, if the shock is close to being invertible – or more generally, recoverable – α^2_{LB} will be close to α^2. As mentioned above, α^2_{LB} will in fact be close to α^2 as long as the y_t process is highly informative about the $\varepsilon_{1,t}$ process at *some* frequency. For example, the observed macro variables y_t may not perfectly reveal the short-run fluctuations of an unobserved technology shock, so recoverability fails ($R^2_\infty < 1$); yet a long-lag two-sided moving average of GDP growth may well approximate the low-frequency cycles of the technology shock. See Section 4 for a concrete example.

Degree of noninvertibility. The identified set for the degree of noninvertibility at horizon ℓ follows directly from the identified set for α^2, since

$$R^2_\ell = 1 - \operatorname{Var}(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell}) = \frac{1}{\alpha^2} \times \operatorname{Var}(E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell})),$$

and the variance on the right-hand side above is point-identified. Now similarly define

$$\tilde{R}^2_\ell \equiv 1 - \frac{\operatorname{Var}(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell})}{\operatorname{Var}(\tilde{z}_t)} = \frac{\operatorname{Var}(E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell}))}{\operatorname{Var}(\tilde{z}_t)}$$

as the (point-identified) R^2 in a population regression of $\tilde{z}_t$ on lags and leads of y_τ up to time $\tau = t + \ell$. Then the identified set for the degree of noninvertibility R^2_ℓ at horizon ℓ equals

$$\Bigg[\underbrace{\tilde{R}^2_\ell}_{\frac{\alpha^2}{\alpha^2+\sigma_v^2}\times R^2_\ell} \; , \; \underbrace{\frac{\operatorname{Var}(\tilde{z}_t)}{2\pi \sup_{\omega\in[0,\pi]} s_{\tilde{z}^\dagger}(\omega)} \times \; \tilde{R}^2_\ell}_{\frac{1}{2\pi \sup_{\omega\in[0,\pi]} s_{\varepsilon_1^\dagger}(\omega)}\times R^2_\ell} \Bigg]. \tag{16}$$

This identified set implies conditions under which the data allow us to either confirm invertibility with certainty or confirm noninvertibility with certainty.

Proposition 2. *Assume $\alpha^2_{LB} > 0$. The identified set for R^2_0 contains 1 if and only if the instrument residual $\tilde{z}_t$ does not Granger cause the macro observables y_t. The identified set for R^2_0 equals the singleton $\{1\}$ if and only if $\tilde{R}^2_0 = 1$.*

According to Proposition 2, we know for sure that $\varepsilon_{1,t}$ is invertible if and only if $\tilde{R}^2_0 = 1$. Similarly, we know for sure that $\varepsilon_{1,t}$ is *non*invertible if and only if $\tilde{z}_t$ Granger causes y_t. As the upper bound in the identified set (16) for R^2_0 shows, consistency with invertibility in our model requires $R^2_0 = 2\pi \sup_{\omega\in[0,\pi]} s_{\varepsilon_1^\dagger}(\omega)$, which can only hold if future values of y_t do not help in predicting the current hidden shock $\varepsilon_{1,t}$. This is the case in the static model, explaining why 1 always lies in the static-only identified set for R^2_0.

Variance decompositions/ratios. We now turn to the identification of variance decompositions, the main parameters of interest. The identified sets for the FVR and FVD defined in Section 2 are different. For the FVR, simply observe that

$$FVR_{i,\ell} = \frac{\sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2}{\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})} = \frac{1}{\alpha^2} \times \frac{\sum_{m=0}^{\ell-1} \operatorname{Cov}(y_{i,t}, \tilde{z}_{t-m})^2}{\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})}.$$

Hence, as in the static case, the identified set for $FVR_{i,\ell}$ equals the identified set for $\frac{1}{\alpha^2}$, scaled by the (point-identified) second fraction on the far right-hand side above.

The identified set for the FVD requires more work. Intuitively, the (point-identified) full forecasting variance $\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})$ conflates pure forecasting uncertainty (which enters the denominator of the FVD) and invertibility-related forecasting uncertainty (which does not). We thus need to bound the contribution of pure forecasting uncertainty. The following proposition summarizes our results.

Proposition 3. *Let there be given a joint spectral density for $w_t = (y_t', \tilde{z}_t)'$ satisfying the assumptions in Proposition 1. Given knowledge of $\alpha \in (\alpha_{LB}, \alpha_{UB}]$, the largest possible value of the forecast variance decomposition $FVD_{i,\ell}$ is 1 (the trivial bound), while the smallest possible value is given by*

$$\frac{\sum_{m=0}^{\ell-1} \operatorname{Cov}(y_{i,t}, \tilde{z}_{t-m})^2}{\sum_{m=0}^{\ell-1} \operatorname{Cov}(y_{i,t}, \tilde{z}_{t-m})^2 + \alpha^2 \operatorname{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\leq t})}. \tag{17}$$

Here $\tilde{y}_t^{(\alpha)} = (\tilde{y}_{1,t}^{(\alpha)}, \dots, \tilde{y}_{n_y,t}^{(\alpha)})'$ denotes a stationary Gaussian time series with spectral density $s_{\tilde{y}^{(\alpha)}}(\omega) = s_y(\omega) - \frac{2\pi}{\alpha^2} s_{y\tilde{z}}(\omega) s_{y\tilde{z}}(\omega)^*$, $\omega \in [0, 2\pi]$. Expression (17) is monotonically decreasing in α, so the overall lower bound on $FVD_{i,\ell}$ is attained by $\alpha = \alpha_{UB}$; in this boundary case we can represent $\tilde{y}_t^{(\alpha_{UB})} = y_t - E(y_t \mid \{\tilde{z}_\tau\}_{-\infty<\tau\leq t})$.

The upper bound on the ℓ-period-ahead FVD is always 1, for any $\ell \geq 1$. This is achieved by a model in which all shocks, except the first one, only affect y_t after an ℓ-period delay.

The expression for the lower bound (17) has a simple interpretation. Even if α is known, the denominator $\operatorname{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty<\tau\leq t})$ of the FVD is not identified due to the lack of information about shocks other than the first. Although we can upper-bound this conditional variance by the denominator of the FVR, this upper bound is not sharp. Instead, to maximize the denominator, as much forecasting noise as possible should be of the pure forecasting variety, and not related to noninvertibility. For all shocks except for $\varepsilon_{1,t}$, this is achievable through a Wold decomposition construction (Hannan, 1970, Thm. 2″, p. 158). Given α, we know the contribution of the first shock to y_t; the residual after removing this contribution has the distribution of $\tilde{y}_t^{(\alpha)}$, as defined in the proposition. If α is not known, the smallest possible value of the lower bound (17) is attained at the largest possible value of α, namely α_{UB}, for which $\varepsilon_{1,t}$ contributes the least to forecasts of y_t.

Sufficient conditions for point identification. Although we have shown that partial identification analysis is informative in the general model, we now give a variety of sufficient conditions that ensure point identification of α and thus the FVR and degree of noninvertibility. We also discuss identification of historical decompositions. Proposition 3 showed that even point identification of α is insufficient to point-identify the FVD, although a sharp and informative lower bound can be computed.[9]

The first set of sufficient conditions relates to the informativeness of the macro aggregates y_t for the hidden shock $\varepsilon_{1,t}$. In this category, our weakest condition for point identification is that the data y_t is perfectly informative about $\varepsilon_{1,t}$ at *some* frequency, i.e., there exists an $\overline{\omega} \in [0, \pi]$ such that $s_{\varepsilon_1^\dagger}(\overline{\omega}) = s_{\varepsilon_1}(\cdot) = \frac{1}{2\pi}$. Then $\alpha = \alpha_{LB}$, so the FVR and degree of noninvertibility are identified. This assumption is not testable. A stronger but more easily interpretable identifying assumption is recoverability, i.e. $\varepsilon_{1,t}^\dagger \equiv E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau<\infty}) = \varepsilon_{1,t}$. This assumption is rejected by the data if the identified set for R^2_∞ does not contain 1.[10] Under recoverability, we have both $\alpha = \alpha_{LB}$ and $\tilde{z}_t^\dagger = \alpha \varepsilon_{1,t}$. Recoverability is a restrictive assumption, but at least it is a meaningfully weaker requirement than invertibility for many economic applications, as discussed in Section 4. Recoverability is implied by the usual SVAR assumption that there are as many shocks as variables, $n_\varepsilon = n_y$.[11] Our analysis thus demonstrates how restrictive the latter assumption really is.[12]

A second set of sufficient conditions for point identification relates to instrument informativeness. If the instrument is perfectly informative, so $\tilde{z}_t = \alpha \varepsilon_{1,t}$, then identification proceeds in accordance with the logic behind local projections (Jordà, 2005). Alternatively, point identification obtains if multiple instruments with mutually independent "first-stage disturbances" are available, cf. Section 3.3.

[9]Stock & Watson (2017) show that the FVD is point-identified if all shocks are assumed invertible.

[10]This is the case if and only if $z_t^\dagger$ is serially correlated at some lag.

[11]If the joint spectrum of y_t and $\tilde{z}_t$ is positive definite at every frequency, then $n_\varepsilon = n_y$ implies that $\Theta(L)^{-1}$ is a well-defined two-sided lag polynomial (Brockwell & Davis, 1991, Thm. 3.1.3), so that $\varepsilon_t = \Theta(L)^{-1} y_t$ and *all* shocks are recoverable.

[12]The proof of Proposition 1 shows that α is partially identified with the same sharp bounds as above even if we know that the number of shocks n_ε can be at most $n_y + 1$. Thus, no identifying power is gained from the knowledge that the number of shocks is "small", unless that means $n_\varepsilon = n_y$.

Under either recoverability or perfect instrument informativeness, we can point-identify the *historical decomposition* corresponding to the identified shock. The historical decomposition of variable $y_{i,t}$ at time t attributable to the identified shock is defined as $\sum_{\ell=0}^{\infty} \Theta_{i,1,\ell}\varepsilon_{1,t-\ell}$. This object is identified because both the impulse responses and the time series of the shock itself are identified, as argued above.

3.3 Extension: multiple instruments

We now argue that identification analysis in the model with multiple IVs for the shock of interest ($n_z \geq 2$) can be reduced to the single-IV setting without loss of generality. If the distribution of the data is consistent with the model, the available IVs can be transformed to a single IV that captures all identifying power for the parameters of interest.

The multiple-IV model is testable, unlike the single-IV model. As in the single-IV case, define the projection residual

$$\tilde{z}_t \equiv z_t - E(z_t \mid \{y_\tau, z_\tau\}_{-\infty<\tau<t}) = \alpha\lambda\varepsilon_{1,t} + \Sigma_v^{1/2} v_t. \tag{18}$$

Appendix A.1 shows that the testable implication of the multiple-IV model is that the cross-spectrum $s_{y\tilde{z}}(\omega)$ has a rank-1 factor structure. The validity of the multiple-IV model can be rejected if and only if this factor structure fails.

When the multiple-IV model is consistent with the distribution of the data, identification analysis can be reduced to the single-IV case in Section 3.2. Specifically, Appendix A.1 shows that (i) λ is point-identified, and (ii) the identified sets for α, variance decompositions, and the degree of noninvertibility are the same as the identified sets that exploit only the scalar instrument

$$\breve{z}_t = \frac{1}{\lambda' \operatorname{Var}(\tilde{z}_t)^{-1}\lambda} \lambda' \operatorname{Var}(\tilde{z}_t)^{-1}\tilde{z}_t. \tag{19}$$

Because $\breve{z}_t$ is a linear combination of all n_z instruments, the identified sets are narrower than if we had used any one instrument $z_{k,t}$ in isolation.[13]

Additional restrictions on the IVs can ensure point identification. In particular, if $n_z \geq 2$ and the researcher is willing to restrict Σ_v to be diagonal, then α is point-identified from any off-diagonal element of $\operatorname{Var}(\tilde{z}_t) = \Sigma_v + \alpha^2\lambda\lambda'$, since λ is point-identified.

[13] Intuitively, $\operatorname{Var}(\varepsilon_{1,t} \mid \tilde{z}_t) = 1 - \alpha^2\lambda' \operatorname{Var}(\tilde{z}_t)^{-1}\lambda = 1 - \alpha^2 \operatorname{Var}(\breve{z}_t)^{-1}$, so the the single instrument $\breve{z}_t$ delivers the same upper bound on α as the vector of instruments $\tilde{z}_t$.

4 Illustration using a structural macro model

We use the workhorse business cycle model of Smets & Wouters (2007) to illustrate the informativeness of our partial identification bounds for the degree of noninvertibility and variance decompositions. We show how the width of the identified sets depends on the strength of the instrument and the informativeness of the macro variables for the unknown shock. The model's monetary policy shock is nearly invertible, so standard SVAR methods would deliver reasonable identification of this shock. In contrast, invertibility is a very poor approximation when identifying the effects of forward guidance (anticipated monetary) shocks or of technology shocks. Nevertheless, our sharp bounds on variance decompositions and the degree of noninvertibility are informative for all three shocks. For clarity, we focus entirely on population bounds in this section, assuming the spectral density of the data is known. The econometrician uses our LP-IV techniques and does not exploit the underlying structure of the model.

Model. We employ the Smets & Wouters (2007) model. Throughout, we parametrize the model according to the posterior mode estimates of Smets & Wouters (2007).[14] Following the empirical literature on monetary policy shock transmission, we assume the econometrician observes aggregate output, inflation, and the short-term policy interest rate. These macro aggregates are all stationary in the model, so they should be viewed as deviations from trend. The model features seven unobserved shocks, so not all shocks can be invertible.

The econometrician observes a single external instrument z_t for the shock of interest $\varepsilon_{1,t}$:

$$z_t = \alpha\varepsilon_{1,t} + \sigma_v v_t.$$

We normalize $\alpha = 1$ throughout and compute identified sets for two different degrees of informativeness of the external instrument, $\frac{1}{1+\sigma_v^2} \in \{0.25, 0.5\}$.[15]

We consider three different shocks of interest: a monetary shock, a forward guidance shock, and a technology shock. The monetary shock is nearly invertible, but the others are not. The forward guidance shock is instead nearly recoverable, whereas only the *long-run* cycles of the technology shock can be accurately recovered from the data. Nevertheless, we

[14]Our implementation of the Smets-Wouters model is based on *Dynare* replication code kindly provided by Johannes Pfeifer. The code is available at `https://sites.google.com/site/pfeiferecon/dynare`.

[15]These correspond to an F statistic of $\frac{1}{3}T$ and T, respectively, in an infeasible regression of z_t on $\varepsilon_{1,t}$ with sample size T. T should be viewed as measured in quarters.

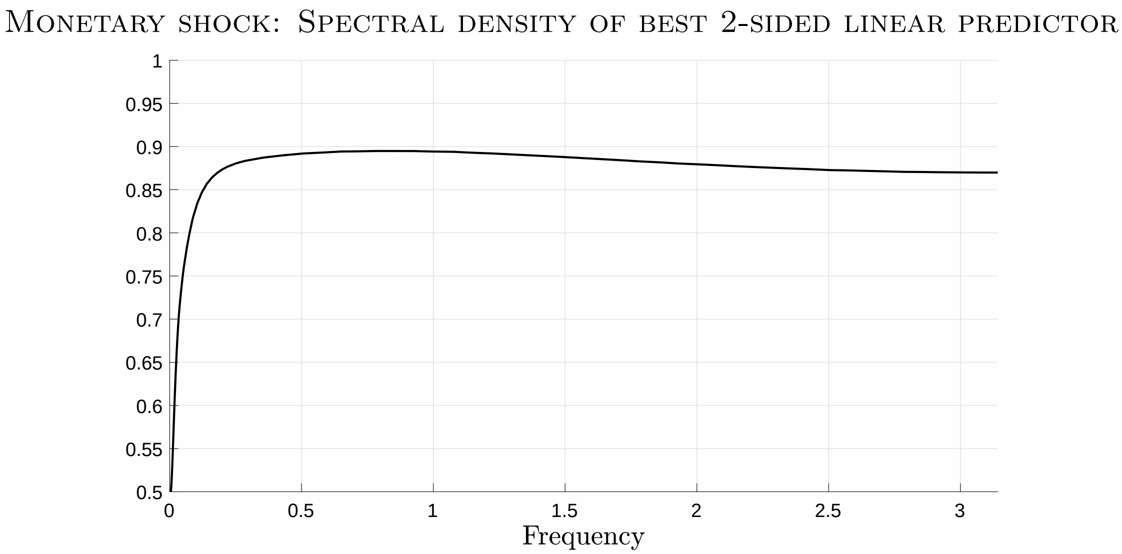

Figure 1: Scaled spectral density $2\pi s_{\varepsilon_1^\dagger}(\cdot)$ of the best two-sided linear predictor of the monetary shock. A frequency ω corresponds to a cycle of length $\frac{2\pi}{\omega}$ quarters.

show that partial identification analysis is informative about the effects of all three shocks.

Monetary shock. We first consider identification of monetary policy shocks. These are defined as shocks to the serially correlated disturbance in the model's Taylor rule.

The monetary shock is nearly invertible in our parametrization. Specifically, the collection of all past and current values of the observable macro variables explain a fraction $R_0^2 = 0.8705$ of the variance of the shock, as already shown by Wolf (2017). The infinite past, present, and *future* of the observables yield only slightly sharper identification, with $R_\infty^2 = 0.8767$. Figure 1 shows the spectral density $s_{\varepsilon_1^\dagger}(\cdot)$ of the two-sided best linear predictor of the monetary shock based on all macro variables. The data are essentially equally informative about medium and high frequencies of the monetary shock, whereas the long-run cycles of the shock cannot be accurately recovered from the data. At the peak of the spectral density, the observables explain a fraction 0.8958 of the variance of that particular cyclical component of the monetary shock; hence, $\alpha_{LB} = \sqrt{0.8958} = 0.9465$, which is close to the truth of 1.

Because the shock is nearly invertible, the upper bounds of the identified sets for the forecast variance ratio and the degree of noninvertibility are close to the truth, while the lower bounds depend on the informativeness of the IV. The identified set for R_ℓ^2 is given by

$$\left[\frac{1}{1+\sigma_v^2} \times R_\ell^2, \quad \frac{1}{0.9465^2} \times R_\ell^2\right].$$

Monetary shock: Identified set of FVRs

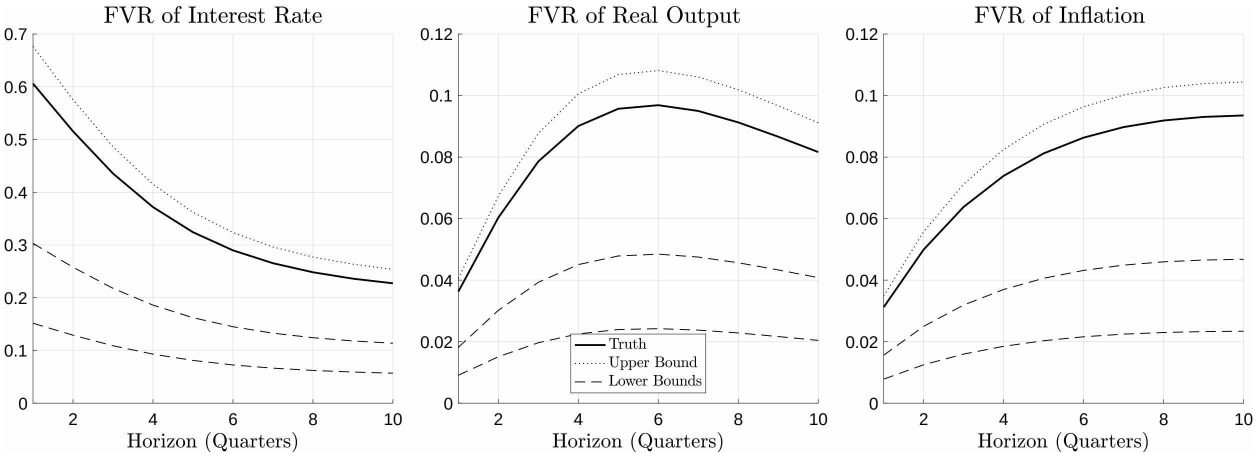

Figure 2: Horizon-by-horizon identified sets for FVRs up to 10 quarters. The two lower bounds are for $\frac{1}{1+\sigma_v^2} = 0.25$ (lower dashed line) and $\frac{1}{1+\sigma_v^2} = 0.5$.

Since $R_0^2, R_\infty^2 < 0.9465^2$, the data reject both invertibility and recoverability. Moreover, the econometrician can rule out that the shock is highly noninvertible, provided the IV is not weak. Figure 2 displays the identified set of the FVR at different forecast horizons.[16] The upper and lower bounds are proportional to the true FVRs. The lower bound scales one-for-one with instrument informativeness, while the upper bound scales one-for-one with the maximal informativeness of the data for the shock across frequencies. The upper bounds are thus close to the true FVRs in this application with a near-invertible shock, whereas the relative informativeness of the lower bounds depends entirely on the strength of the IV.

For FVDs, the lower bound of the identified set also depends on the informativeness of the IV, while the upper bound always equals the trivial value 1. Figure 3 depicts the identified sets for FVDs, ommitting the trivial upper bound. The lower bound is now not simply proportional to the true FVD, due to the intricacies of bounding the denominator of the FVD. In this application, the lower bound is nevertheless approximately equal to the true FVD scaled by instrument informativeness $\frac{1}{1+\sigma_v^2}$.

Due to the near-invertibility of the monetary shocks, incorrectly imposing invertibility (or recoverability) would cause the researcher to overstate the forecast variance ratios and historical decompositions by a modest 14 per cent each. As shown above, the data are in fact sufficiently informative so as to reject both the invertibility and recoverability assumptions. If

[16] Throughout this paper, the identified sets for FVRs are constructed horizon by horizon. However, the *joint* uncertainty about FVRs at different horizons is caused by uncertainty about the single parameter α.

MONETARY SHOCK: IDENTIFIED SET OF FVDS

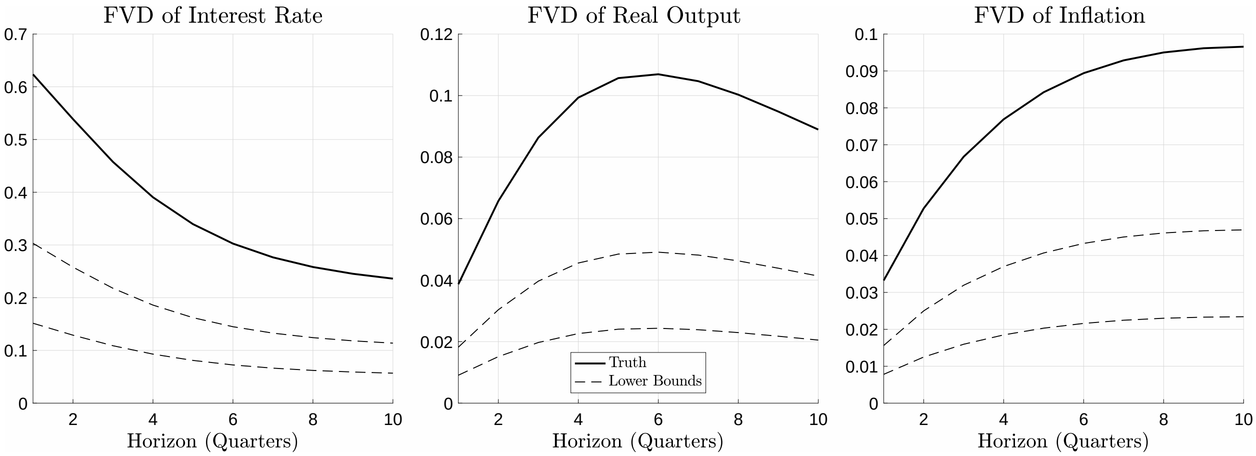

Figure 3: Horizon-by-horizon identified sets for FVDs up to 10 quarters. The two lower bounds are for $\frac{1}{1+\sigma_v^2} = 0.25$ (lower dashed line) and $\frac{1}{1+\sigma_v^2} = 0.5$. Upper bound not shown.

a researcher instead (also incorrectly) imposes the weaker and untestable assumption that the data is perfectly informative about the shock at *some* frequency, then all the aforementioned objects of interest would be overstated by around 11 per cent ($\frac{1}{\alpha_{LB}^2} \approx 1.11$).

FORWARD GUIDANCE SHOCK. We now augment the model to include forward guidance shocks, a type of news shock. A forward guidance shock is identical to a monetary shock, except it is anticipated two quarters in advance by economic agents.[17]

As is common with news shocks, the forward guidance shock is highly noninvertible but approximately recoverable. The wedge between information contained in the infinite past and information contained in the entire time series of observables is sizable: Contemporaneous informativeness is limited, with $R_0^2 = 0.0792$, but looking two quarters ahead basically returns us to the level of informativeness for the standard monetary shock, with $R_2^2 = 0.8731$ and $R_\infty^2 = 0.8813$. Intuitively, the decisions of forward-looking agents today reveal a nonzero but limited amount of information about the shock that is about to materialize two quarters from now. As the shock then hits, aggregates respond strongly, whereupon the econometrician learns as much about the shock as she did in the benchmark model with a non-anticipated shock. Thus, with news shocks, the incremental bite of two-sided analysis can be substantial.[18]

[17]Formally, we implement forward guidance by changing the baseline Smets & Wouters (2007) model so that the monetary shock has time subscript $t-2$ instead of t.

[18]In news shock models with as many shocks as observables – e.g., the fiscal foresight models of Leeper

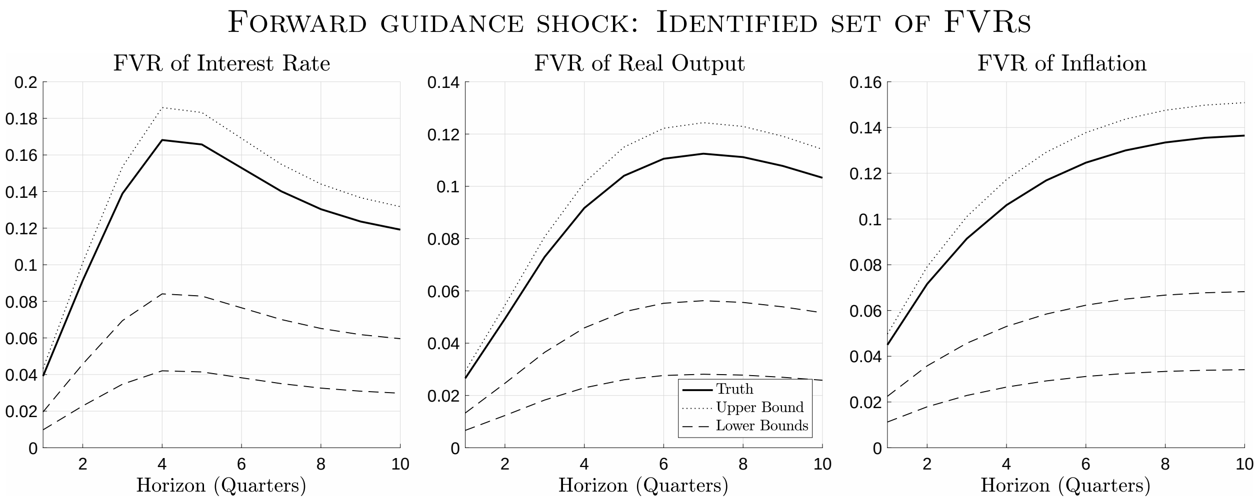

Figure 4: Horizon-by-horizon identified sets for FVRs up to 10 quarters. The two lower bounds are for $\frac{1}{1+\sigma_v^2} = 0.25$ (lower dashed line) and $\frac{1}{1+\sigma_v^2} = 0.5$.

Despite the high degree of noninvertibility, the identified sets for the FVRs of the forward guidance shock are as informative as those for the monetary shock, as shown in Figure 4.[19] This demonstrates that our partial identification analysis is not only robust to noninvertibility – its quantitative usefulness does not depend on the degree of noninvertibility *per se*. In stark contrast, identification that incorrectly imposes invertibility (e.g., SVARs) would overstate variance decompositions by a factor of $1/0.0792 \approx 13$ (!).[20] Recoverability-based identification would err by a more modest factor of $1/0.8813 \approx 1.13$.

Technology shock. Finally, we consider identification of technology shocks, defined as an innovation to the autoregressive process of total factor productivity.

Unlike the monetary and forward guidance shocks, the technology shock is far from recoverable; nevertheless, our bounds remain informative. In the model, the technology shock is much more important in accounting for low-frequency cycles of the data than it is for high-frequency cycles. The technology shock is far from being recoverable, let alone

et al. (2013) – the strict recoverability assumption is satisfied. Invertibility-based identification is biased unless y_t contains variables that respond immediately when news arrive.

[19]Note that the FVR of the interest rate is much lower than for the standard monetary shock (Figure 2). The interest rate now *drops* when news arrive, and overall it moves less because the endogenous rule-based response of the monetary authority counteracts the effects of the news prior to the policy impulse.

[20]To be exact, standard SVAR-IV methods would overstate *impact* impulse responses by a factor of $1/\sqrt{0.0792} \approx 3.6$ and so impact variance decompositions by a factor of 13. Subsequent impulse responses would not be proportional to true responses, due to the imposed VAR dynamics (Stock & Watson, 2017).

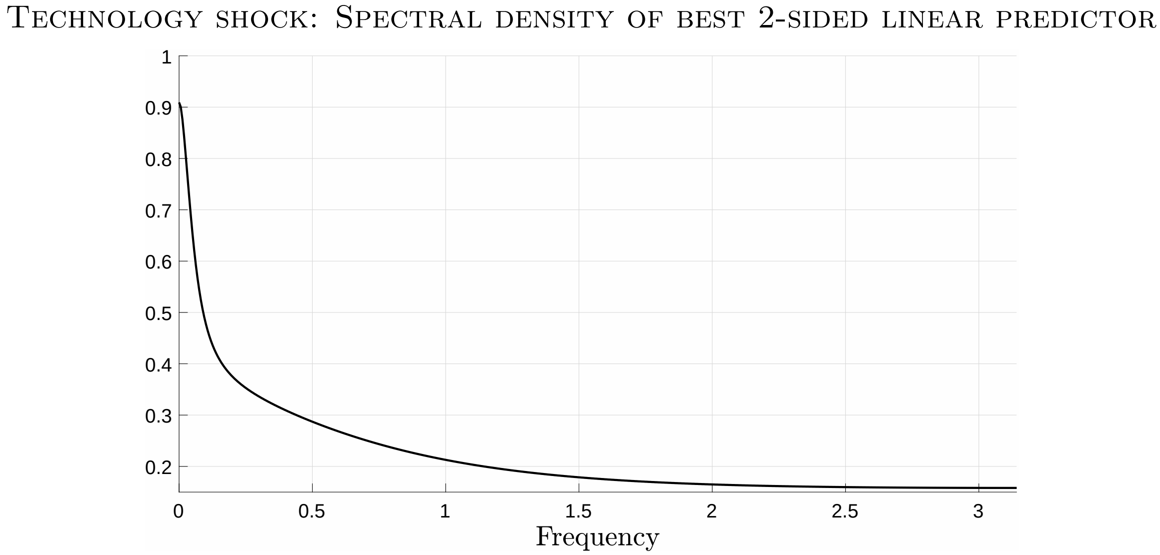

Figure 5: Scaled spectral density $2\pi s_{\varepsilon_1^\dagger}(\cdot)$ of the best two-sided linear predictor of the technology shock. A frequency ω corresponds to a cycle of length $\frac{2\pi}{\omega}$ quarters.

invertible, with $R_0^2 = 0.2007$ and $R_\infty^2 = 0.2209$. However, the data are very informative about the lowest-frequency cycles of the technology shock, as shown in Figure 5. As a result, $\alpha_{LB}^2 = 0.9092$ is close to the true value of 1, and the upper bounds of our identified sets for FVRs and the degree of noninvertibility (not shown) yield tight identification. In contrast, identification that incorrectly imposes either invertibility or recoverability of the shock overstates the FVR by a factor of about 5.

5 Inference

To make the identification analysis practically useful, we develop partial identification robust confidence intervals and tests. In a first step, the researcher estimates a reduced-form VAR model, which is then used in a second step to derive sample analogues of our population bounds. Using the general partial identification confidence procedures of Imbens & Manski (2004) and Stoye (2009), we construct confidence intervals for both the parameters and for the identified sets; a Bayesian implementation is available for the latter. Our confidence intervals have asymptotically valid frequentist coverage. We also discuss a test for invertibility.

We assume the availability of a single instrument z_t for notational simplicity. The generalization to multiple instruments is straight-forward, as discussed in Section 3.3.

Reduced-form VAR. In this section we assume that the second-moment properties of the data are captured by a reduced-form VAR in $(y_t', z_t)'$ with known, finite lag length p.

Assumption 1. *There exist $(n_y+1)\times(n_y+1)$ matrices A_ℓ, $\ell = 1, 2, \dots, p$, and a symmetric positive definite $(n_y+1)\times(n_y+1)$ matrix Σ such that the spectral density of $W_t = (y_t', z_t)'$ is given by*

$$s_W(\omega) = \left(I_{n_y+1} - \sum_{\ell=1}^{p} A_\ell e^{-i\omega\ell}\right)^{-1} \Sigma \left(I_{n_y+1} - \sum_{\ell=1}^{p} A_\ell e^{-i\omega\ell}\right)^{-1*}, \quad \omega \in [0, 2\pi],$$

and such that all roots of the polynomial $x \mapsto \det(I_{n_y+1} - \sum_{\ell=1}^{p} A_\ell x^\ell)$ are outside the unit circle. Moreover, there exist estimators $\hat{A}_1, \dots, \hat{A}_\ell, \hat{\Sigma}$ (measurable functions of the data $W_1, \dots, W_T$) such that $\hat{\vartheta} \equiv (\text{vec}(\hat{A}_1)', \dots, \text{vec}(\hat{A}_p)', \text{vech}(\hat{\Sigma})')'$ is a (pointwise) asymptotically normal estimator of $\vartheta \equiv (\text{vec}(A_1)', \dots, \text{vec}(A_p)', \text{vech}(\Sigma)')'$:

$$\sqrt{T}(\hat{\vartheta} - \vartheta) \xrightarrow{d} N(0, \Omega) \quad as \quad T \to \infty,$$

where Ω is positive definite. Finally, there exists an estimator $\hat{\Omega}$ which is (pointwise) consistent for Ω.

We assume a reduced-form VAR structure for four reasons. First, VARs are known to be able to approximate any spectral density function arbitrarily well as the VAR lag length tends to infinity. Second, familiar estimators $\hat{\vartheta}$ and $\hat{\Omega}$ of the parameters of the spectrum and the asymptotic variance are available; for example, the usual least-squares VAR toolkit will do, provided the data is stationary (Lütkepohl, 2005, Ch. 3). Third, the VAR structure facilitates the development of a test of invertibility and of Bayesian inference procedures. Fourth, VAR-based inference amounts to redoing our population calculations from Section 3 on a spectrum of a particular functional form (namely a VAR spectrum with the particular estimated parameters $\hat{\vartheta}$). All inequalities satisfied in the population must then also hold in any finite sample, thus guaranteeing nonempty identified sets, for example (up to numerical error, but not statistical error). This property is harder to achieve with a nonparametric kernel smoothing estimator of the spectrum, say.

The assumed finite-lag reduced-form VAR structure is restrictive but a reasonable starting point. In practice, we suggest estimating the lag length p by information criteria or likelihood ratio tests (Lütkepohl, 2005, Ch. 4). In ongoing work, we are developing asymptotic results that allow $p \to \infty$, thus making our inference procedures truly nonparametric. We emphasize, though, that assuming a finite-lag reduced-form VAR is less restrictive than doing SVAR-IV inference: We do not assume that the reduced-form VAR residuals span the true structural shocks. For example, we continue to allow the number of structural shocks to possibly exceed the number of variables in the VAR.

Invertibility test. It is straight-forward to test for invertibility of the shock of interest using the estimated reduced-form VAR. We showed in Proposition 2 that the data is consistent with invertibility of $\varepsilon_{1,t}$ if and only if $\tilde{z}_t$ does not Granger cause y_t. Granger non-casuality of $\tilde{z}_t$ for y_t is equivalent with Granger non-causality of z_t for y_t. Under Assumption 1, it is well known that a test of the Granger non-causality null hypothesis amounts to a test of the exclusion restrictions that lags of z_t do not enter the reduced-form VAR equations for y_t (Lütkepohl, 2005, Ch. 2.3 and 3.6). This test has power against all Granger causal alternatives, so it has power against all noninvertible alternatives by Proposition 2.[21]

Confidence intervals. We now construct partial identification robust confidence intervals for identified sets and for the true parameters. Here we rely heavily on the inference methods pioneered by Imbens & Manski (2004) and refined by Stoye (2009).

We start out by defining notation. Under Assumption 1, all identified sets derived in Section 3.2 are of the form $[\underline{h}(\vartheta), \overline{h}(\vartheta)]$, where $\underline{h}(\cdot)$ and $\overline{h}(\cdot)$ are continuous functions mapping the VAR parameter space into the real line, and such that $\underline{h}(\cdot) \leq \overline{h}(\cdot)$. A (pointwise) consistent estimator of the identified set $[\underline{h}(\vartheta), \overline{h}(\vartheta)]$ is then given by the plug-in interval

$$[\underline{h}(\hat{\vartheta})\, ,\ \overline{h}(\hat{\vartheta})].$$

Let $\hat{\Delta} \equiv \overline{h}(\hat{\vartheta}) - \underline{h}(\hat{\vartheta})$ denote the width of the estimate of the identified set. Assume $\underline{h}(\cdot)$ and $\overline{h}(\cdot)$ are continuously differentiable at the true VAR parameters ϑ with $1 \times \dim(\vartheta)$ dimensional Jacobian functions $\dot{\underline{h}}(\cdot)$ and $\dot{\overline{h}}(\cdot)$. Define the standard errors of $\underline{h}(\hat{\vartheta})$ and $\overline{h}(\hat{\vartheta})$,

$$\hat{\underline{\sigma}} \equiv \sqrt{T^{-1}\dot{\underline{h}}(\hat{\vartheta})\hat{\Omega}\dot{\underline{h}}(\hat{\vartheta})'}\, , \quad \hat{\overline{\sigma}} \equiv \sqrt{T^{-1}\dot{\overline{h}}(\hat{\vartheta})\hat{\Omega}\dot{\overline{h}}(\hat{\vartheta})'},$$

and their correlation,

$$\hat{\rho} \equiv \frac{T^{-1}\dot{\underline{h}}(\hat{\vartheta})\hat{\Omega}\dot{\overline{h}}(\hat{\vartheta})'}{\hat{\underline{\sigma}} \times \hat{\overline{\sigma}}}.$$

[21] Stock & Watson (2017) develop an LP-IV invertibility test which they argue is different from a Granger causality test. Under our assumptions, Proposition 2 implies that an invertibility test that controls size and has nontrivial power against all noninvertible alternatives must be equivalent with a Granger causality test.

Finally, let $\Phi(\cdot)$ denote the standard normal cumulative distribution function.

We now construct a confidence interval for the entire identified set. The interval

$$\left[\underline{h}(\hat{\vartheta}) - \Phi^{-1}(1-\beta/2)\hat{\underline{\sigma}}\,,\ \overline{h}(\hat{\vartheta}) + \Phi^{-1}(1-\beta/2)\hat{\overline{\sigma}}\right] \tag{20}$$

is a (pointwise) asymptotically valid level-$(1-\beta)$ confidence interval for the identified set $[\underline{h}(\vartheta), \overline{h}(\vartheta)]$. That is, the above interval contains the *entire* identified set in at least $100(1-\beta)\%$ of repeated experiments, asymptotically. This follows from the delta method and the arguments of Imbens & Manski (2004).

Next, we construct a confidence interval for the true parameter of interest. By definition of the identified set, the true parameter is contained in $[\underline{h}(\vartheta), \overline{h}(\vartheta)]$, but we know nothing else about the true parameter. Although the interval (20) trivially has asymptotic coverage of at least $1-\beta$ for the true parameter, Imbens & Manski (2004) showed that it is possible to develop a narrower interval with the same property. As in Stoye (2009, p. 1305), define the two scalars $\hat{\underline{c}}, \hat{\overline{c}}$ as the minimizers of the objective function

$$\hat{\underline{\sigma}} \times \hat{\underline{c}} + \hat{\overline{\sigma}} \times \hat{\overline{c}},$$

subject to the two constraints

$$\Pr\left(-\hat{\underline{c}} \leq U_1,\ \hat{\rho} U_1 \leq \hat{\overline{c}} + \frac{\hat{\Delta}}{\hat{\overline{\sigma}}} + \sqrt{1-\hat{\rho}^2} \times U_2\right) \geq 1-\beta,$$

$$\Pr\left(-\hat{\underline{c}} - \frac{\hat{\Delta}}{\hat{\underline{\sigma}}} - \sqrt{1-\hat{\rho}^2} \times U_2 \leq \hat{\rho} U_1,\ U_1 \leq \hat{\overline{c}}\right) \geq 1-\beta.$$

Here the probabilities are taken solely over the distribution of $(U_1, U_2)'$, which is bivariate standard normal. The above minimization problem is easy to solve numerically, cf. Stoye (2009, Appendix B). Given these definitions, the interval

$$\left[\underline{h}(\hat{\vartheta}) - \hat{\underline{c}} \times \hat{\underline{\sigma}}\,,\ \overline{h}(\hat{\vartheta}) + \hat{\overline{c}} \times \hat{\overline{\sigma}}\right]$$

is a (pointwise) asymptotically valid level-$(1-\beta)$ confidence interval for the true parameter. Again, this result follows from the delta method and the results in Stoye (2009), who builds on Imbens & Manski (2004).

To implement the above confidence interval procedures, the researcher needs to compute the VAR estimator $\hat{\vartheta}$, the asymptotic variance matrix estimate $\hat{\Omega}$, the bound estimates $\underline{h}(\hat{\vartheta})$

and $\overline{h}(\hat{\vartheta})$, and the derivatives of the bounds $\dot{\underline{h}}(\hat{\vartheta})$ and $\dot{\overline{h}}(\hat{\vartheta})$. Appendix A.2 provides formulas for the bounds and derivatives in terms of the VAR parameters. Simple bootstrap and Bayesian implementations are also available, see below.

We now discuss how to resolve the complication that the upper bound of the identified sets for $\frac{1}{\alpha^2}$, R_0^2, and the FVR may not be continuously differentiable in the VAR parameters. The issue arises because α_{LB}^2 is given by the maximum of a certain function, and when this function has multiple maxima at the true VAR parameters (e.g., when the spectral density of $\tilde{z}_t^\dagger$ is flat, as in the recoverable case), continuous differentiability of α_{LB}^2 in the VAR parameters ϑ may fail (Gafarov et al., 2017). In this case, delta method inference will be unreliable. As a remedy, we suggest replacing the maximum $\alpha_{LB}^2 = 2\pi \sup_{\omega \in [0,\pi]} s_{\tilde{z}^\dagger}(\omega)$ in all our bounds with the smaller average value $\text{Var}(\tilde{z}_t^\dagger) = \int_0^{2\pi} s_{\tilde{z}^\dagger}(\omega)\, d\omega$, cf. the inequality (15). The latter object is continuously differentiable in the VAR parameters, so inference using the above methods is unproblematic. Use of the non-sharp bound does lead to a power loss, but the loss is small if the shock $\varepsilon_{1,t}$ is close to being recoverable, or if the informational content of the data for the shock does not vary substantially across frequencies, as explained in Sections 3.2 and 4.[22] Note that continuous differentiability of the bounds for the FVD obtains without modifications.

Our confidence intervals are pointwise valid in both senses of the word. First, we focus on constructing a confidence interval for each parameter of interest separately, as opposed to capturing the joint uncertainty of several parameters at once. It is an interesting topic for future research to develop simultaneous confidence bands for, say, the FVD across forecast horizons. Second, our asymptotics are pointwise in the true parameters; we do not derive the coverage under the worst-case data generating process.[23] In particular, we ignore finite-sample issues caused by weak instruments, i.e., $\alpha_{LB} \approx 0$.

[22] More generally, we can lower-bound α_{LB}^2 by $\int_0^{2\pi} r(\omega) s_{\tilde{z}^\dagger}(\omega)\, d\omega$, where $r(\cdot)$ is a nonnegative function such that $\int_0^{2\pi} r(\omega)\, d\omega = 2\pi$. If the researcher has prior information about the frequencies ω at which y_t is particularly informative about $\varepsilon_{1,t}$, then $r(\omega)$ can be chosen to weight these frequencies more heavily. This yields a more informative bound than $\text{Var}(\tilde{z}_t^\dagger)$, while preserving continuous differentiability.

[23] The Imbens & Manski (2004) and Stoye (2009) procedures are designed to control coverage uniformly over the width of the identified set. We do not discuss uniform asymptotics here because if we wanted to assume that the convergence assumptions in Assumption 1 hold uniformly, it would be necessary to assume that the magnitude of the largest eigenvalue of the VAR polynomial is bounded away from 1, in which case the width of the identified set (for all our objects of interest) would also be bounded away from zero. Hence, in this case, the uniform asymptotic validity of the confidence procedures is a trivial matter.

Bootstrap implementation. The calculation of derivatives in the confidence interval formulas above is obviated by the bootstrap. Suppose we have a method for bootstrapping the estimator $\hat{\vartheta}$ (Kilian & Lütkepohl, 2017, Ch. 12). Then we can compute $\underline{\hat{\sigma}}$ as the bootstrap standard deviation of $\underline{h}(\hat{\vartheta})$, $\hat{\overline{\sigma}}$ as the bootstrap standard deviation of $\overline{h}(\hat{\vartheta})$, and $\hat{\rho}$ as the bootstrap correlation of $\underline{h}(\hat{\vartheta})$ and $\overline{h}(\hat{\vartheta})$. By plugging into the same confidence interval formulas as above, we achieve the same (pointwise) asymptotic coverage probability as the delta method confidence intervals, provided that Assumption 1 and an appropriate bootstrap consistency condition hold.

Bayesian implementation. Finally, we discuss a Bayesian credible interval for the identified set. Bayesian inference on the identified set can be motivated by a desire for structural inference to be robust to the choice of prior on structural parameters, conditional on a fixed prior on reduced-form parameters (Giacomini & Kitagawa, 2015). Suppose we form a prior for the reduced-form VAR parameters ϑ, and we have a method for drawing from the posterior distribution of ϑ (e.g., Kilian & Lütkepohl, 2017, Ch. 5). Then we can compute the interval $[E(\underline{h}(\vartheta) \mid \text{data}),\ E(\overline{h}(\vartheta) \mid \text{data})]$, a Bayesian estimate of the identified set. Moreover, we can compute the shortest interval which covers at least $100(1-\beta)\%$ of the posterior draws of the intervals $[\underline{h}(\vartheta), \overline{h}(\vartheta)]$. This smallest interval is then a probability-$(1-\beta)$ credible interval for the identified set (Giacomini & Kitagawa, 2015).

A Supplementary results

This appendix elaborates on the theoretical identification analysis and inference procedures. First, we show that the multiple-IV model can be reduced to the single-IV model. Second, we provide formulas to implement the confidence intervals.

A.1 Multiple instruments

Here we show that the multiple-IV model is testable, but if it is consistent with the data, then identification analysis can be reduced to the single-IV case.

Define the IV residual vector $\tilde{z}_t$ as in equation (18). The multiple-IV model (1)–(3) implies the following cross-spectrum between y_t and $\tilde{z}_t$:

$$s_{y\tilde{z}}(\omega) = \frac{\alpha}{2\pi}\Theta(e^{-i\omega})e_1\lambda', \quad \omega \in [0, 2\pi]. \tag{21}$$

Thus, the cross-spectrum has rank-1 factor structure: It equals a nonconstant column vector times a constant row vector. This testable property turns out to be exactly what characterizes the multiple-IV model.

Proposition 4. *Let a spectrum $s_w(\omega)$ for $w_t = (y_t', \tilde{z}_t')'$ be given, satisfying the assumptions of Proposition 1. There exists a model of the form (1), (3), and (18) which generates the spectrum $s_w(\omega)$ if and only if there exist n_y-dimensional real vectors ζ_ℓ, $\ell \geq 0$, and an n_z-dimensional constant real vector η of unit length such that*

$$s_{y\tilde{z}}(\omega) = \zeta(e^{-i\omega})\eta', \quad \omega \in [0, 2\pi], \tag{22}$$

where $\zeta(L) = \sum_{\ell=0}^{\infty}\zeta_\ell L^\ell$.

Assuming henceforth that the factor structure obtains, we now show that identification in the multiple-IV model reduces to the single-IV case. It is convenient first to reparametrize the model slightly, by setting $\Sigma_v = \Sigma_{\tilde{z}} - \alpha^2\lambda\lambda'$ and treating $\Sigma_{\tilde{z}}$ as a basic model parameter instead of Σ_v. We then impose the requirement that $\Sigma_{\tilde{z}} - \alpha^2\lambda\lambda'$ be positive semidefinite. Clearly, $\Sigma_{\tilde{z}} = \text{Var}(\tilde{z}_t)$ is point-identified. Next, note from (21) that λ is point-identified and equal to the η vector in equation (22). This is because any rank-1 factorization of a matrix is identified up to sign and scale, and we have normalized η to have length 1. Let Ξ be any

$(n_z - 1) \times n_z$ matrix such that $\Xi \Sigma_{\tilde{z}}^{-1/2} \lambda = 0$. Define the $n_z \times n_z$ matrix

$$Q \equiv \begin{pmatrix} \frac{1}{\lambda' \Sigma_{\tilde{z}}^{-1} \lambda} \lambda' \Sigma_{\tilde{z}}^{-1} \\ \Xi \Sigma_{\tilde{z}}^{-1/2} \end{pmatrix}.$$

Since Q is point-identified (given a choice of Ξ), it is without loss of generality to perform identification analysis based on the linearly transformed IV residuals

$$Q\tilde{z}_t = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_{1,t} + \tilde{v}_t, \quad \tilde{v}_t \sim N\left(0, \begin{pmatrix} \frac{1}{\lambda' \Sigma_{\tilde{z}}^{-1} \lambda} - \alpha^2 & 0 \\ 0 & \Xi\Xi' \end{pmatrix}\right).$$

Notice, however, that α only enters into the equation for the first element of $Q\tilde{z}_t$, and the $(n_z - 1)$ last elements of $Q\tilde{z}_t$ are independent of the first element (and independent of y_t at all leads and lags). Hence, it is without loss of generality to limit attention to the first element of $Q\tilde{z}_t$ when performing identification analysis for the impulse responses $\Theta_{i,j,\ell}$ and the scale parameter α. The first element of $Q\tilde{z}_t$ equals $\breve{z}_t$ as defined in equation (19) in the main text.[24]

A.2 Formulas for implementing the confidence intervals

Here we provide formulas needed to construct the partial identification robust confidence intervals in Section 5. Specifically, assuming the spectrum of $(y_t', z_t)'$ has VAR structure as in Assumption 1, we state formulas for the interval bounds in terms of the reduced-form VAR coefficients ϑ, and we discuss how to compute the derivatives of these expressions.

We introduce the notation

$$A(L) = I_{n_y+1} - \sum_{\ell=1}^{p} A_\ell L^\ell, \quad \Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_z \end{pmatrix},$$

where the upper left block in the partition is $n_y \times n_y$. Let J denote the top n_y rows of I_{n_y+1}, and let e_{n_y+1} denote the $(n_y + 1)$-dimensional unit vector with 1 as the last element. Define

[24] The above display implies that we must have $\alpha^2 \leq (\lambda' \operatorname{Var}(\tilde{z}_t)^{-1} \lambda)^{-1}$, which is precisely what the upper bound for α^2 yields when applied to $\breve{z}_t$.

also the matrix functions

$$s_y(\omega; A, \Sigma) = \frac{1}{2\pi} J A(e^{-i\omega})^{-1} \Sigma A(e^{-i\omega})^{-1*} J', \quad \omega \in [0, 2\pi],$$

$$s_{y\tilde{z}}(\omega; A, \Sigma) = \frac{1}{2\pi} J A(e^{-i\omega})^{-1} \Sigma e_{n_y+1}, \quad \omega \in [0, 2\pi].$$

Bounds for α. To compute the bounds for α, observe that

$$\alpha_{UB}^2 = \text{Var}(\tilde{z}_t) = \Sigma_z,$$

$$\alpha_{LB}^2 = 2\pi \max_{\omega \in [0,\pi]} s_{y\tilde{z}}(\omega; A, \Sigma)^* s_y(\omega; A, \Sigma)^{-1} s_{y\tilde{z}}(\omega; A, \Sigma),$$

$$\text{Var}(\tilde{z}_t^\dagger) = 2 \int_0^\pi s_{y\tilde{z}}(\omega; A, \Sigma)^* s_y(\omega; A, \Sigma)^{-1} s_{y\tilde{z}}(\omega; A, \Sigma)\, d\omega,$$

all of which can be easily computed, at least numerically. The derivative of α_{UB}^2 with respect to the VAR parameters is obvious, the derivative of α_{LB}^2 can be obtained by the envelope theorem if the maximum is uniquely attained, and the derivative of $\text{Var}(\tilde{z}_t^\dagger)$ can be obtained by differentiating under the integral sign.

Bounds for R_0^2. The only missing ingredient to computing the identified set for the degree of noninvertibility is $\text{Var}(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t})$. We can approximate this quantity arbitrarily well as $M \to \infty$ by

$$\text{Var}(\tilde{z}_t \mid \{y_\tau\}_{t-M\leq\tau\leq t}) = \text{Var}(\tilde{z}_t) - (\Sigma_{zy}, 0_{1\times n_y M}) V_M(A, \Sigma)^{-1} (\Sigma_{zy}, 0_{1\times n_y M})',$$

where $V_M(A, \Sigma)$ is the usual variance-covariance matrix of $(y_t', y_{t-1}', \ldots, y_{t-M}')'$ implied by the VAR, see for example Lütkepohl (2005, Ch. 2.1.1). Derivatives can be computed by the chain rule, finite differences, or automatic differentiation. It is advisable to check robustness with respect to the choice of M.

Bounds for FVR. To compute the identified set for $FVR_{i,\ell}$, we need

$$\text{Cov}(y_{i,t}, \tilde{z}_{t-m}) = \int_0^{2\pi} e^{i\omega m} s_{y\tilde{z}}(\omega; A, \Sigma)\, d\omega, \quad m = 0, 1, \ldots, \ell - 1,$$

which can be computed by numerical integration; the derivative is obtained by differentiating under the integral sign. We also need $\text{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})$. This object is well approximated for large M by[25]

$$\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{t-M\leq\tau\leq t}) = \operatorname{Var}(y_{i,t}) - (\operatorname{Cov}(y_{i,t+\ell}, y_t), \dots, \operatorname{Cov}(y_{i,t+\ell}, y_{t-M}))V_M(A,\Sigma)^{-1} \times (\operatorname{Cov}(y_{i,t+\ell}, y_t), \dots, \operatorname{Cov}(y_{i,t+\ell}, y_{t-M}))',$$

where $V_M(A,\Sigma)$ was defined above. All objects on the right-hand side can be computed using standard VAR formulas (Lütkepohl, 2005, Ch. 2.1.1), and derivatives can be computed by the chain rule, finite differences, or automatic differentiation.

Bounds for FVD. To compute the overall lower bound for the FVD, we need $\operatorname{Var}(\tilde{y}_{i,t+\ell}^{(\alpha_{UB})} \mid \{\tilde{y}_\tau^{(\alpha_{UB})}\}_{-\infty<\tau\leq t})$. As before, we approximate this by $\operatorname{Var}(\tilde{y}_{i,t+\ell}^{(\alpha_{UB})} \mid \{\tilde{y}_\tau^{(\alpha_{UB})}\}_{t-M\leq\tau\leq t})$ for large M. The same formula used above for $\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{t-M\leq\tau\leq t})$ applies, where covariances are obtained from

$$\operatorname{Cov}(\tilde{y}_{t+\ell}^{(\alpha_{UB})}, \tilde{y}_t^{(\alpha_{UB})}) = \operatorname{Cov}(y_{t+\ell}, y_t) - \frac{1}{\alpha_{UB}^2}\sum_{m=0}^{\infty} \operatorname{Cov}(y_t, \tilde{z}_{t-m-\ell}) \operatorname{Cov}(y_t, \tilde{z}_{t-m})'.$$

The sum can be truncated when the contribution of additional terms is very small.

[25] A more numerically stable strategy is to compute $\operatorname{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{t-M\leq\tau\leq t})$ using the Kalman filter. Since the conditional variance does not depend on the realized values of the conditioning variables, the "data" fed into the Kalman filter can be all zeros.

B Proofs and auxiliary lemmas

B.1 Auxiliary lemmas

Lemma 1. *Let B be an $n \times n$ Hermitian positive definite complex-valued matrix and b an n-dimensional complex-valued column vector. Let x be a nonnegative real scalar. Then $B - x^{-1}bb^*$ is positive (semi)definite if and only if $x > (\geq)\, b^* B^{-1} b$.*

Proof. We focus on the semidefiniteness statement. Decompose $B = B^{1/2}B^{1/2*}$ and define $\tilde{b} = B^{-1/2}b$. The statement of the lemma is equivalent with the statement that $I_n - x^{-1}\tilde{b}\tilde{b}^*$ is positive semidefinite if and only if $x \geq b^*b$. Let ν be an arbitrary n-dimensional complex vector satisfying $\nu^*\nu = 1$. Then

$$\nu^* \left(I_n - x^{-1}\tilde{b}\tilde{b}^* \right) \nu = 1 - \frac{\tilde{b}^*\tilde{b}}{x} \cos^2 \left(\theta(\nu, \tilde{b}) \right),$$

where $\theta(\nu, \tilde{b})$ is the angle between ν and $\tilde{b}$. Evidently, $x^{-1}\tilde{b}^*\tilde{b} \leq 1$ is precisely the condition needed to ensure that the above display is nonnegative for every choice of ν. □

Lemma 2. *Let x_t and $\tilde{x}_t$ be two stationary n-dimensional Gaussian time series whose spectral densities $s_x(\omega)$ and $s_{\tilde{x}}(\omega)$ are such that $s_{\tilde{x}}(\omega) - s_x(\omega)$ is positive semidefinite for all $\omega \in [0, 2\pi]$. Then $\mathrm{Var}(\mu' x_{t+\ell} \mid \{x_\tau\}_{-\infty<\tau\leq t}) \leq \mathrm{Var}(\mu' \tilde{x}_{t+\ell} \mid \{\tilde{x}_\tau\}_{-\infty<\tau\leq t})$ for all $\ell = 1, 2, \ldots$ and all constant vectors $\mu \in \mathbb{R}^n$.*

Proof. We may define an n-dimensional stationary Gaussian process ν_t with spectral density $s_\nu(\omega) = s_{\tilde{x}}(\omega) - s_x(\omega)$, $\omega \in [0, 2\pi]$, and such that the ν_t process is independent of the x_t process. Then the process $\check{x}_t = x_t + \nu_t$ has the same distribution as the $\tilde{x}_t$ process. Hence,

$$\begin{aligned}
\mathrm{Var}(\mu' \tilde{x}_{t+\ell} \mid \{\tilde{x}_\tau\}_{-\infty<\tau\leq t}) &= \mathrm{Var}(\mu' \check{x}_{t+\ell} \mid \{\check{x}_\tau\}_{-\infty<\tau\leq t}) \\
&\geq \mathrm{Var}(\mu' \check{x}_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty<\tau\leq t}) \\
&= \mathrm{Var}(\mu' x_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty<\tau\leq t}) + \mathrm{Var}(\mu' \nu_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty<\tau\leq t}) \\
&\geq \mathrm{Var}(\mu' x_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty<\tau\leq t}) \\
&= \mathrm{Var}(\mu' x_{t+\ell} \mid \{x_\tau\}_{-\infty<\tau\leq t}).
\end{aligned}$$

The second equality above uses that the independence of the x_t and ν_t processes implies that $x_{t+\ell}$ and $\nu_{t+\ell}$ are independent also conditional on $\{x_\tau, \nu_t\}_{-\infty<\tau\leq t}$. □

B.2 Proof of Proposition 1

Let α and the spectrum $s_w(\omega)$ be given. Define the n_y-dimensional vectors

$$\overline{\Theta}_{\bullet,1,\ell} = \alpha^{-1}\operatorname{Cov}(y_t, \tilde{z}_{t-\ell}), \quad \ell \geq 0,$$

and the corresponding vector lag polynomial

$$\overline{\Theta}_{\bullet,1}(L) = \sum_{\ell=0}^{\infty} \overline{\Theta}_{\bullet,1,\ell} L^{\ell}.$$

Since $\alpha^2 \leq \alpha_{UB}^2$, we may define $\overline{\sigma}_v = \sqrt{\operatorname{Var}(\tilde{z}_t) - \alpha^2}$. Since $\alpha^2 > \alpha_{LB}^2$, Lemma 1 implies that

$$s_y(\omega) - \frac{2\pi}{\alpha^2} s_{y\tilde{z}}(\omega) s_{y\tilde{z}}(\omega)^* = s_y(\omega) - \frac{1}{2\pi}\overline{\Theta}_{\bullet,1}(e^{-i\omega})\overline{\Theta}_{\bullet,1}(e^{-i\omega})^*$$

is positive definite for every $\omega \in [0, 2\pi]$. Hence, the Wold decomposition theorem (Hannan, 1970, Thm. 2″, p. 158) implies that there exists an $n_y \times n_y$ matrix lag polynomial $\tilde{\Theta}(L) = \sum_{\ell=0}^{\infty} \tilde{\Theta}_{\ell} L^{\ell}$ such that[26]

$$s_y(\omega) - \frac{1}{2\pi}\overline{\Theta}_{\bullet,1}(e^{-i\omega})\overline{\Theta}_{\bullet,1}(e^{-i\omega})^* = \frac{1}{2\pi}\tilde{\Theta}(e^{-i\omega})\tilde{\Theta}(e^{-i\omega})^*, \quad \omega \in [0, 2\pi].$$

Thus, the following model for $w_t = (y_t', \tilde{z}_t)'$ generates the desired spectrum $s_w(\omega)$:

$$\begin{aligned} y_t &= \overline{\Theta}_{\bullet,1}(L)\overline{\varepsilon}_{1,t} + \tilde{\Theta}(L)\tilde{\varepsilon}_t, \\ \tilde{z}_t &= \alpha\overline{\varepsilon}_{1,t} + \overline{\sigma}_v \overline{v}_t, \\ (\overline{\varepsilon}_{1,t}, \tilde{\varepsilon}_t', \overline{v}_t)' &\overset{i.i.d.}{\sim} N(0, I_{n_y+2}). \end{aligned}$$

As an aside, note that the construction requires only $n_\varepsilon = n_y + 1$ structural shocks, $\overline{\varepsilon}_{1,t} \in \mathbb{R}$ and $\tilde{\varepsilon}_t \in \mathbb{R}^{n_y}$. □

B.3 Proof of Proposition 2

The second statement of the proposition is immediate. We now prove the first statement.

If the identified set contains 1, then there must exist an $\overline{\alpha} \in [\alpha_{LB}, \alpha_{UB}]$ and i.i.d.,

[26] We can rule out a deterministic term in the Wold decomposition because a continuous and positive definite spectral density satisfies the full-rank condition of Hannan (1970, p. 162).

independent standard Gaussian processes $\bar{\varepsilon}_{1,t}$ and $\bar{v}_t$ such that (i) $\tilde{z}_t = \bar{\alpha} \times \bar{\varepsilon}_{1,t} + \bar{v}_t$, (ii) $\bar{v}_t$ is uncorrelated with y_t at all leads and lags, and (iii) $\bar{\varepsilon}_{1,t}$ lies in the closed linear span of $\{y_\tau\}_{-\infty<\tau\leq t}$. This immediately implies the "only if" part of the corollary.

For the "if" part, assume $\tilde{z}_t$ does not Granger cause y_t. By the equivalence of Sims and Granger causality, $\tilde{z}_t^\dagger = E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau<\infty}) = E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t})$. Note that the latter best linear predictor is white noise since, for any $\ell \geq 1$,

$$\begin{aligned}\operatorname{Cov}\left(E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t}), y_{t-\ell}\right) &= \operatorname{Cov}(\tilde{z}_t, y_{t-\ell}) - \operatorname{Cov}\left(\tilde{z}_t - E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t}), y_{t-\ell}\right) \\ &= 0 - 0,\end{aligned}$$

using the fact that $\tilde{z}_t$ is a projection residual. In conclusion, the best linear predictor $\tilde{z}_t^\dagger$ of $\tilde{z}_t$ given $\{y_\tau\}_{-\infty<\tau<\infty}$ depends only on $\{y_\tau\}_{-\infty<\tau\leq t}$ and it has a constant spectrum. From the expression for α_{LB}^2, we get that $\alpha_{LB}^2 = \operatorname{Var}(E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau\leq t}))$, which further yields $\alpha_{LB}^2 = \operatorname{Var}(\tilde{z}_t)\tilde{R}_0^2$. Hence, expression (16) implies that the upper bound of the identified set for R_0^2 equals 1. □

B.4 Proof of Proposition 3

The proof proceeds in two steps. First, for a given known α, we show that $FVD_{i,\ell}$ is sharply bounded above by 1 and below by (17). Second, we show that the lower bound is monotonically decreasing in α, so that the overall lower bound is attained by α_{UB}.

1. Given $\alpha \in (\alpha_{LB}, \alpha_{UB}]$, the numerator of $FVD_{i,\ell}$ is point-identified (see below), so we need only concern ourselves with the denominator. We can write the denominator as

$$\begin{aligned}\operatorname{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty<\tau\leq t}) &= \sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2 + \sum_{j=2}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2 \\ &= \frac{1}{\alpha^2} \sum_{m=0}^{\ell-1} \operatorname{Cov}(y_{i,t}, \tilde{z}_{t-m})^2 + \sum_{j=2}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2. \qquad (23)\end{aligned}$$

   Given α, the first term in (23) is point-identified (note that it equals the numerator of the FVD), while the second is not. To upper-bound $FVD_{i,\ell}$, we seek to make that second term as small as possible. In fact, we can always set it to 0. To see this, let $\{\Theta_{\bullet,j,m}\}_{2\leq j\leq n_\varepsilon, 0\leq m<\infty}$ denote some sequence of impulse responses for the structural shocks $j \neq 1$ that is consistent with the second-moment properties of the data. Since $\alpha \in (\alpha_{LB}, \alpha_{UB}]$, such a sequence exists by Proposition 1. Now, for a given forecast horizon ℓ, instead consider the new sequence $\{\breve{\Theta}_{\bullet,j,m}\}_{2\le j\le n_\varepsilon, 0\le m<\infty}$, defined via

$$\breve{\Theta}_{\bullet,j,m} = \begin{cases} 0_{n_y\times 1} & \text{if } m \le \ell - 1, \\ \Theta_{\bullet,j,m-\ell} & \text{if } m > \ell - 1. \end{cases}$$

Then the stochastic process induced by $\{\breve{\Theta}_{\bullet,j,m}\}_{2\le j\le n_\varepsilon, 0\le m<\infty}$ has the exact same second-moment properties as the (by assumption admissible) stochastic process induced by $\{\Theta_{\bullet,j,m}\}_{2\le j\le n_\varepsilon, 0\le m<\infty}$. However, by construction, we now have $FVD_{i,\ell} = 1$, as claimed.

For the lower bound, we want to make the second term in (23) as large as possible. Given a known $\alpha \in (\alpha_{LB}, \alpha_{UB}]$, define

$$\tilde{y}_t^{(\alpha)} = (\tilde{y}_{1,t}^{(\alpha)}, \dots, \tilde{y}_{n_y,t}^{(\alpha)})' \equiv y_t - \frac{1}{\alpha}\sum_{\ell=0}^{\infty} \text{Cov}(y_t, \tilde{z}_{t-\ell})\varepsilon_{1,t-\ell} = \sum_{j=2}^{n_\varepsilon}\sum_{\ell=0}^{\infty} \Theta_{\bullet,j,\ell}\varepsilon_{j,t-\ell},$$

whose spectral density is given by the expression stated in the proposition. We have

$$\text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t}) \ge \text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\varepsilon_{j,\tau}\}_{2\le j\le n_\varepsilon, -\infty<\tau\le t}) = \sum_{j=2}^{n_\varepsilon}\sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2,$$

so the second term in (23) has an point-identified upper bound. Thus, given α, $FVD_{i,\ell}$ is bounded below by the expression (17).

We now argue that the lower bound (17) is attained by an admissible model with the given α. To that end, consider the Wold decomposition of $\tilde{y}_t^{(\alpha)} = \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \tilde{\varepsilon}_{t-\ell}$, where the $\tilde{\Theta}_\ell$ matrices are $n_y \times n_y$, and $\tilde{\varepsilon}_t$ is n_y-dimensional i.i.d. standard normal and spanned by $\{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t}$.[27] Then $\text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t}) = \sum_{j=2}^{n_\varepsilon}\sum_{m=0}^{\ell-1} \tilde{\Theta}_{i,j,m}^2$, so the following model attains the lower bound (17) and is consistent with the given spectrum $s_w(\cdot)$:

$$\begin{aligned} y_t &= \frac{1}{\alpha}\sum_{\ell=0}^{\infty} \text{Cov}(y_t, \tilde{z}_{t-\ell})\bar{\varepsilon}_{1,t} + \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \tilde{\varepsilon}_{t-\ell}, \\ \tilde{z}_t &= \alpha\bar{\varepsilon}_{1,t} + \sqrt{\text{Var}(\tilde{z}_t) - \alpha^2} \times \bar{v}_t, \qquad (24) \\ (\bar{\varepsilon}_{1,t}, \tilde{\varepsilon}_t', \bar{v}_t)' &\overset{i.i.d.}{\sim} N(0, I_{n_y+2}). \end{aligned}$$

2. Lemma 2 implies that $\text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t})$ is increasing in α. Hence, the expression (17) is decreasing in α, as claimed. At $\alpha = \alpha_{UB}$, the representation (24) has $\tilde{z}_t = \alpha_{UB}\bar{\varepsilon}_{1,t}$,

[27] Since $\alpha > \alpha_{LB}$, the Wold decomposition has no deterministic term, cf. the proof of Proposition 1.

so we can represent $\tilde{y}_t^{(\alpha_{UB})} = y_t - E(y_t \mid \{\bar{\varepsilon}_{1,\tau}\}_{-\infty<\tau\leq t}) = y_t - E(y_t \mid \{\tilde{z}_\tau\}_{-\infty<\tau\leq t})$. □

B.5 Proof of Proposition 4

The "only if" part was proved already in the text of Appendix A.1. For the "if" part, assume that the cross-spectrum has the given factor structure. Since $\tilde{z}_t$ is serially uncorrelated, we can write $s_{\tilde{z}}(\cdot) = s_{\tilde{z}}$. Because $s_w(\omega)$ is positive definite, the Schur complement

$$s_{\tilde{z}} - s_{y\tilde{z}}(\omega)^* s_y(\omega)^{-1} s_{y\tilde{z}}(\omega) = s_{\tilde{z}} - \eta\zeta(\omega)^* s_y(\omega)^{-1}\zeta(\omega)\eta'$$

is also positive definite. Pre-multiplying the above expression by $\eta' s_{\tilde{z}}^{-1}$, post-multiplying by $s_{\tilde{z}}^{-1}\eta$, and rearranging the positive definiteness condition, we obtain the implication that

$$2\pi\zeta(\omega)^* s_y(\omega)^{-1}\zeta(\omega) < \frac{2\pi}{\eta' s_{\tilde{z}}^{-1}\eta}, \quad \omega \in [0, 2\pi].$$

Now choose any $\bar{\alpha} \geq 0$ such that $\bar{\alpha}^2$ lies strictly between the left- and right-hand sides in the above inequality. The matrix

$$\bar{\Sigma}_v \equiv 2\pi s_{\tilde{z}} - \bar{\alpha}^2 \eta\eta'$$

is then positive definite by Lemma 1. Moreover, the same lemma implies that

$$s_y(\omega) - \frac{2\pi}{\bar{\alpha}^2}\zeta(\omega)\zeta(\omega)^*$$

is positive definite for all $\omega \in [0, 2\pi]$. If we set $\bar{\Theta}_{\bullet,1}(L) = (2\pi/\bar{\alpha})\zeta(L)$, the same arguments as in the proof of Proposition 1 show that there exists an $n_y \times n_y$ matrix polynomial $\tilde{\Theta}(L)$ such that the following model achieves the desired spectrum $s_w(\omega)$:

$$y_t = \bar{\Theta}_{\bullet,1}(L)\bar{\varepsilon}_{1,t} + \tilde{\Theta}(L)\tilde{\varepsilon}_t,$$
$$\tilde{z}_t = \bar{\alpha}\eta\bar{\varepsilon}_{1,t} + \bar{\Sigma}_v^{1/2}\bar{v}_t,$$
$$(\bar{\varepsilon}_{1,t}, \tilde{\varepsilon}_t', \bar{v}_t')' \overset{i.i.d.}{\sim} N(0, I_{n_y+n_z+1}).$$

Note that η assumes the role of λ. □

References

Angrist, J. D., Jordà, Ò., & Kuersteiner, G. M. (2017). Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited. *Journal of Business & Economic Statistics*. Forthcoming.

Barnichon, R. & Brownlees, C. (2017). Impulse Response Estimation by Smooth Local Projections. Manuscript, Universitat Pompeu Fabra.

Barnichon, R. & Matthes, C. (2017). Functional Approximations of Impulse Responses (FAIR): New Insights into the Asymmetric Effects of Monetary Policy. Manuscript, Federal Reserve Bank of San Francisco.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer Series in Statistics. Springer.

Caldara, D. & Kamps, C. (2017). The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers. *The Review of Economic Studies*, *84*(3), 1015–1040.

Den Haan, W. J. & Levin, A. (1997). A practitioner's guide to robust covariance matrix estimation. In G. Maddala & C. Rao (Eds.), *Handbook of Statistics*, volume 15 chapter 12, (pp. 299–342). North-Holland.

Gafarov, B., Meier, M., & Montiel Olea, J. L. (2017). Delta-Method Inference for a Class of Set-Identified SVARs. Manuscript, Columbia University.

Gertler, M. & Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76.

Giacomini, R. & Kitagawa, T. (2015). Robust inference about partially identified SVARs. Manuscript, University College London.

Hannan, E. (1970). *Multiple Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Imbens, G. W. & Manski, C. F. (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica*, *72*(6), 1845–1857.

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, *95*(1), 161–182.

Jordà, Ò., Schularick, M., & Taylor, A. M. (2017). Large and state-dependent effects of quasi-random monetary experiments. CEPR Discussion Paper DP11801.

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

Leeper, E. M., Walker, T. B., & Yang, S.-C. S. (2013). Fiscal Foresight and Information Flows. *Econometrica*, *81*(3), 1115–1145.

Lippi, M. & Reichlin, L. (1994). VAR analysis, nonfundamental representations, Blaschke matrices. *Journal of Econometrics*, *63*(1), 307–325.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Mertens, K. (2015). Advances in Empirical Macroeconomics, Lecture 2. Lecture slides, Bonn Summer School.

Mertens, K. & Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, *103*(4), 1212–1247.

Moon, H. R., Schorfheide, F., & Granziera, E. (2013). Inference for VARs Identified with Sign Restrictions. Manuscript, University of Pennsylvania.

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

Ramey, V. & Zubairy, S. (2017). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. *Journal of Political Economy*. Forthcoming.

Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of Macroeconomics*, volume 2 chapter 2, (pp. 71–162). Elsevier.

Sims, C. A. & Zha, T. (2006). Does Monetary Policy Generate Recessions? *Macroeconomic Dynamics*, *10*(02), 231–272.

Smets, F. & Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, *97*(3), 586–606.

Stock, J. H. (2008). What's New in Econometrics: Time Series, Lecture 7. Lecture slides, NBER Summer Institute.

Stock, J. H. & Watson, M. W. (2012). Disentangling the Channels of the 2007–09 Recession. *Brookings Papers on Economic Activity*, *2012*(1), 81–135.

Stock, J. H. & Watson, M. W. (2016). Factor Models and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of Macroeconomics*, volume 2 chapter 8, (pp. 415–525). Elsevier.

Stock, J. H. & Watson, M. W. (2017). Identification and Estimation of Dynamic Causal Effects in Macroeconomics. Manuscript based on 2017 Sargan Lecture, Royal Economic Society.

Stoye, J. (2009). More on Confidence Intervals for Partially Identified Parameters. *Econometrica*, *77*(4), 1299–1315.

Watson, M. W. (1993). Measures of Fit for Calibrated Models. *Journal of Political Economy*, *101*(6), 1011–1041.

Wolf, C. K. (2017). Masquerading Shocks in Sign-Restricted VARs. Manuscript, Princeton University.