Local Projections and VARs Estimate the Same Impulse Responses*

Mikkel Plagborg-Møller
Princeton University

Christian K. Wolf
Princeton University

This version: June 26, 2019
First version: November 16, 2018

Abstract: We prove that local projections and Vector Autoregressions (VARs) estimate the same impulse response functions. This nonparametric result only requires the lag structures in the two specifications to be unrestricted. We discuss several implications: (i) Local projection and VAR estimators should not be thought of as conceptually separate procedures; instead, they belong to a spectrum of dimension reduction techniques that share the same estimand but have different finite-sample bias-variance properties. (ii) VAR-based structural estimation can equivalently be performed using local projections, and *vice versa*. (iii) Valid structural estimation with an instrument (also known as a proxy variable) can be carried out by ordering the instrument first in a recursive VAR, even if the shock of interest is noninvertible. (iv) Linear local projections are not more robust to non-linearities than linear VARs.

Keywords: external instrument, impulse response function, local projection, proxy variable, structural vector autoregression. *JEL codes:* C32, C36.

1 Introduction

Modern dynamic macroeconomics studies the propagation of structural shocks (Frisch, 1933; Ramey, 2016). Central to this impulse-propagation paradigm are impulse responses func-

*Email: mikkelpm@princeton.edu and ckwolf@princeton.edu. We received helpful comments from Domenico Giannone, Marek Jarociński, Òscar Jordà, Peter Karadi, Lutz Kilian, Dake Li, Pepe Montiel Olea, Valerie Ramey, Giovanni Ricco, Chris Sims, Jim Stock, Mark Watson, and seminar participants at various venues. The first draft of this paper was written while Wolf was visiting the European Central Bank, whose hospitality is gratefully acknowledged.

tions – the dynamic response of a macro aggregate to a structural shock. Following Sims (1980), Structural Vector Autoregression (SVAR) analysis remains the most popular empirical approach to impulse response estimation. Over the past decade, however, starting with Jordà (2005), local projections (LPs) have become an increasingly widespread alternative econometric approach. Unfortunately, so far there exists little guidance as to which method is preferable in empirical practice. Applied to the same estimation problem, LP and VAR-based approaches often give very different answers (Ramey, 2016), and existing simulation evidence on their relative merits is conflicting (Meier, 2005; Kilian & Kim, 2011; Brugnolini, 2018; Nakamura & Steinsson, 2018; Choi & Chudik, 2019).

The central result of this paper is that linear local projections and VARs in fact estimate the exact same impulse responses in population. Specifically, any LP impulse response function can be obtained through an appropriately ordered recursive VAR, and any (possibly non-recursive) VAR impulse response function can be obtained through a LP with appropriate control variables. This *nonparametric* result only requires the lag structures in the two specifications to be unrestricted. Intuitively, a VAR model with sufficiently large lag length captures all covariance properties of the data. Hence, iterated VAR(∞) forecasts coincide with direct LP forecasts. Since impulse responses are just forecasts, LP and VAR impulse response estimands coincide in population. Furthermore, we prove that if only a fixed number p of lags are included in the LP and VAR, then the two impulse response estimands still agree out to horizon p (but not further), again without imposing any parametric assumptions on the data generating process. The equivalence also holds in sample: LP and VAR impulse response estimators coincide asymptotically if the lag lengths in the two specifications tend to infinity with the sample size. In summary, if VAR and LP results differ in population or in sample, it is due to extraneous restrictions on the lag structure.

The nonparametric equivalence of VAR and LP estimands has several implications for structural estimation in applied macroeconometrics.

First, LP and VAR estimators are not conceptually different methods; instead, they are best viewed as sharing the same estimand but lying on opposite ends of a spectrum of finite-sample bias-variance choices for linear projection. Standard LPs effectively provide no dimensionality reduction, while conventional low-order VARs extrapolate shock propagation from the first few autocorrelations of the data. The relative mean-square error of the two methods – and of intermediate dimension reduction techniques, such as shrinkage – necessarily depends on assumptions about the data generating process (DGP). VAR estimators are optimal if the true DGP is exactly a finite-order VAR, but this is rarely the case in theory or practice. In general, no single estimation method dominates, and in principle neither low-dimensional VARs nor low-dimensional LPs should be treated as having special status. The formal equivalence of LP and VAR estimation to direct and iterated forecasting, respectively, means that extant results on mean-square error rankings from the forecasting literature are also applicable to *structural* macroeconometrics (Schorfheide, 2005; Marcellino et al., 2006).

Second, structural estimation with VARs can equally well be carried out using LPs, and *vice versa*. Structural identification – which is a population concept – is logically distinct from the choice of finite-sample dimension reduction technique. We give several examples of classical "SVAR" identification schemes that are easily implemented using local projection techniques, including recursive, long-run, and sign identification. Ultimately, LP-based structural estimation can succeed if and only if SVAR estimation can succeed.

Third, valid structural estimation with an instrument (IV, also known as a proxy variable) can be carried out by ordering the IV first in a recursive VAR à la Ramey (2011). This is because the LP-IV estimand of Stock & Watson (2018) can equivalently be obtained from a recursive (i.e., Cholesky) VAR that contains the IV. Importantly, the "internal instruments" strategy of ordering the IV first in a VAR yields valid impulse response estimates even if the shock of interest is noninvertible, unlike the well-known "external instruments" SVAR-IV approach (Stock, 2008; Stock & Watson, 2012; Mertens & Ravn, 2013). In particular, this result goes through even if the IV is contaminated with measurement error that is unrelated to the shock of interest. In contemporaneous work, Noh (2018) derives a closely related result; our formulation offers additional insights by tying the result to the general equivalence between LPs and VARs.

Fourth, linear local projections are exactly as "robust to non-linearities" in the DGP as VARs. We show that their common estimand may be formally interpreted as a best linear approximation to the underlying, perhaps non-linear, data generating process.

While the existing literature has pointed out connections between LPs and VARs, our contribution is to formally establish a nonparametric equivalence result and derive implications for estimation efficiency and structural identification. Jordà (2005) and Kilian & Lütkepohl (2017, Ch. 12.8) show that, under the assumption of a finite-order VAR model, VAR impulse responses can be estimated consistently through LPs. In this context, Kilian & Lütkepohl also discuss the relative efficiency of the two estimation methods and mention the literature on direct versus iterated forecasts. In contrast, our equivalence result is *non-parametric*, and we further demonstrate how structural VAR orderings map into particular choices of LP control variables, and *vice versa*. Moreover, to our knowledge, our results on long-run/sign identification, LP-IV, and best linear approximations have no obvious parallels in the preceding literature.[1]

Before presenting our results, we need to mention a few caveats. First, we are primarily interested in identification of impulse responses to a single shock, rather than system identification. Second, we focus on impulse responses, not variance decompositions or historical decompositions (see Plagborg-Møller & Wolf, 2019, for identification of these objects). Third, we only explore linear estimators. The equivalence of VAR and LP estimators does not apply if we augment the regressions with nonlinear terms, or if we are directly interested in non-linearities stemming from stochastic volatility, say. Fourth, we restrict attention to stationary time series and do not consider issues related to near-unit roots or cointegration. Fifth, we only discuss the population properties of IV estimators and thus do not consider weak IV issues. Finally, we do not comment on the ease of calculating standard errors for different estimation methods, or on generalizations of these methods outside the aggregate time series context.[2]

Outline. Section 2 presents our core result on the population equivalence of local projections and VARs. Finite-sample estimation is discussed in Section 3, while Section 4 traces out implications for *structural* estimation. We illustrate our equivalence results with a practical application to IV-based identification of monetary policy shocks in Section 5. Section 6 concludes with several recommendations for empirical practice. Some proofs and supplementary results are relegated to Appendix A.

2 Equivalence between Local Projections and VARs

This section presents our core result: Local projections and VARs estimate the same impulse response functions in population. First we establish that local projections are equivalent with recursively identified VARs when the lag structure is unrestricted. Then we extend the

[1] Kilian & Lütkepohl (2017, Ch. 12.8) present alternative arguments for why it is a mistake to assert that finite-order LPs are generally more "robust to model misspecification" than finite-order VAR estimators. They do not appeal to the nonparametric equivalence of the LP and VAR estimands, however.

[2] As is well known, low-order VAR and LP methods present somewhat different issues from the point of view of *inference*. Frequentist and Bayesian inference in VARs is straight-forward (under the assumption that the VAR bias is negligible). LP estimation requires Heteroskedasticity and Autocorrelation Robust inference, and the lack of an explicit low-dimensional generating model makes Bayesian inference challenging. On the other hand, LP inference can be easily applied to panel data settings using off-the-shelf software.

argument to (i) non-recursive identification and (ii) finite lag orders. Finally, we illustrate the results graphically. Our analysis in this section does not assume any specific underlying structural model; we merely exploit properties of linear projections of stationary time series.

2.1 Main result

Suppose the researcher observes data $w_t = (r_t', x_t, y_t, q_t')'$, where r_t and q_t are, respectively, $n_r \times 1$ and $n_q \times 1$ vectors of time series, while x_t and y_t are scalar time series. We are interested in the dynamic response of y_t after an impulse in x_t. The vector time series r_t and q_t (which may each be empty) will serve as control variables. The distinction between them relates to whether they appear as *contemporaneous* controls or not, as will become clear in equations (1) and (2) below.

For now, we only make the following nonparametric assumption.

Assumption 1. *The data $\{w_t\}$ are covariance stationary and purely non-deterministic, with an everywhere nonsingular spectral density matrix. To simplify notation, we proceed as if $\{w_t\}$ were a (strictly stationary) jointly Gaussian vector time series.*

In particular, we assume nothing about the underlying causal structure of the economy, as this section is concerned solely with properties of linear projections. The Gaussianity assumption is made purely for notational simplicity, as this allows us to write conditional expectations instead of linear projections. If we drop the Gaussianity assumption, all calculations below hold with projections in place of conditional expectations.[3]

We will show that, in population, the following two approaches estimate the *same* impulse response function of y_t with respect to an innovation in x_t.

1. Local projection. Consider for each $h = 0, 1, 2, \ldots$ the linear projection

$$y_{t+h} = \mu_h + \beta_h x_t + \gamma_h' r_t + \sum_{\ell=1}^{\infty} \delta_{h,\ell}' w_{t-\ell} + \xi_{h,t}, \tag{1}$$

where $\xi_{h,t}$ is the projection residual, and $\mu_h, \beta_h, \gamma_h, \delta_{h,1}, \delta_{h,2}, \ldots$ the projection coefficients. The *LP impulse response function* of y_t with respect to x_t is given by $\{\beta_h\}_{h\geq 0}$. Notice that the projection (1) controls for the contemporaneous value of r_t but not of q_t.

[3]Throughout we write any linear projection on the span of infinitely many variables as an infinite sum. This is justified under Assumption 1 if the Wold representation has absolutely summable coefficients, since we can then invert it to obtain a VAR(∞) representation.

2. VAR. Consider the multivariate linear "VAR(∞)" projection

$$w_t = c + \sum_{\ell=1}^{\infty} A_\ell w_{t-\ell} + u_t, \tag{2}$$

where $u_t \equiv w_t - E(w_t \mid \{w_\tau\}_{-\infty<\tau<t})$ is the projection residual, and $c, A_1, A_2, \dots$ the projection coefficients. Let $\Sigma_u \equiv E(u_t u_t')$, and define the Cholesky decomposition $\Sigma_u = BB'$, where B is lower triangular with positive diagonal entries. Consider the corresponding recursive SVAR representation

$$A(L) w_t = c + B\eta_t,$$

where $A(L) \equiv I - \sum_{\ell=1}^{\infty} A_\ell L^\ell$ and $\eta_t \equiv B^{-1} u_t$. Notice that r_t is ordered first in the VAR, while q_t is ordered last. Define the lag polynomial

$$\sum_{\ell=0}^{\infty} C_\ell L^\ell = C(L) \equiv A(L)^{-1}.$$

The *VAR impulse response function* of y_t with respect to an innovation in x_t is given by $\{\theta_h\}_{h\geq 0}$, where

$$\theta_h \equiv C_{n_r+2,\bullet,h} B_{\bullet,n_r+1},$$

since x_t and y_t are the (n_r+1)-th and (n_r+2)-th elements in w_t. The notation $C_{i,\bullet,h}$, say, means the i-th row of matrix C_h, while $B_{\bullet,j}$ is the j-th column of matrix B.

Note that our definitions of the LP and VAR estimands include infinitely many lags of w_t in the relevant projections. We consider the case of finitely many lags in Section 2.3, while all finite-sample considerations are relegated to Section 3.

Although LP and VAR approaches are often viewed as conceptually distinct in the literature, they in fact estimate the same population impulse response function.

Proposition 1. *Under Assumption 1, the LP and VAR impulse response functions are equal, up to a constant of proportionality:* $\theta_h = \sqrt{E(\tilde{x}_t^2)} \times \beta_h$ *for all* $h = 0, 1, 2, \dots$*, where* $\tilde{x}_t \equiv x_t - E(x_t \mid r_t, \{w_\tau\}_{-\infty<\tau<t})$.

That is, any LP impulse response function can equivalently be obtained as an appropriately ordered recursive VAR impulse response function. Conversely, any recursive VAR impulse response function can be obtained through a LP with appropriate control variables. We comment on non-recursive identification schemes below. The constant of proportionality in the proposition depends on neither the impulse response horizon h nor on the response variable y_t. The reason for the presence of this constant of proportionality is that the implicit LP innovation $\tilde{x}_t$, after controlling for the other right-hand side variables, does not have variance 1. If we scale the innovation $\tilde{x}_t$ to have variance 1, or if we consider relative impulse responses θ_h/θ_0 (as further discussed below), the LP and VAR impulse response functions coincide.

The intuition behind the result is that a VAR(p) model with $p \to \infty$ is sufficiently flexible that it perfectly captures all covariance properties of the data. Thus, iterated forecasts based on the VAR coincide perfectly with direct forecasts $E[w_{t+h} \mid w_t, w_{t-1}, \dots]$.

Proof. The proof of the proposition relies only on least-squares projection algebra. First consider the LP estimand. By the Frisch-Waugh theorem, we have that

$$\beta_h = \frac{\text{Cov}(y_{t+h}, \tilde{x}_t)}{E(\tilde{x}_t^2)}. \tag{3}$$

For the VAR estimand, note that $C(L) = A(L)^{-1}$ collects the coefficient matrices in the Wold decomposition

$$w_t = \chi + C(L)u_t = \chi + \sum_{\ell=0}^{\infty} C_\ell B \eta_t, \quad \chi \equiv C(1)c.$$

As a result, the VAR impulse responses equal

$$\theta_h = C_{n_r+2,\bullet,h} B_{\bullet,n_r+1} = \text{Cov}(y_{t+h}, \eta_{x,t}), \tag{4}$$

where we partition $\eta_t = (\eta_{r,t}', \eta_{x,t}, \eta_{y,t}, \eta_{q,t}')'$ the same way as $w_t = (r_t', x_t, y_t, q_t')'$. By $u_t = B\eta_t$ and the properties of the Cholesky decomposition, we have[4]

$$\eta_{x,t} = \frac{1}{\sqrt{E(\tilde{u}_{x,t}^2)}} \times \tilde{u}_{x,t}, \tag{5}$$

[4] B is lower triangular, so the (n_r+1)-th equation in the system $B\eta_t = u_t$ is $B_{n_r+1,1:n_r}\eta_{r,t} + B_{n_r+1,n_r+1}\eta_{x,t} = u_{x,t}$, with obvious notation. Since $\eta_{x,t}$ and $\eta_{r,t}$ are uncorrelated, we find $B_{n_r+1,n_r+1}\eta_{x,t} = u_{x,t} - E(u_{x,t} \mid \eta_{r,t}) = u_{x,t} - E(u_{x,t} \mid u_{r,t}) = \tilde{u}_{x,t}$. Expression (5) then follows from $E(\eta_{x,t}^2) = 1$.

where we partition $u_t = (u'_{r,t}, u_{x,t}, u_{y,t}, u'_{q,t})'$ and define[5]

$$\tilde{u}_{x,t} \equiv u_{x,t} - E(u_{x,t} \mid u_{r,t}) = \tilde{x}_t. \tag{6}$$

From (4), (5), and (6) we conclude that

$$\theta_h = \frac{\text{Cov}(y_{t+h}, \tilde{x}_t)}{\sqrt{E(\tilde{x}_t^2)}},$$

and the proposition now follows by comparing with (3). □

In conclusion, LPs and VARs offer two equivalent ways of arriving at the same population parameter (3), up to a scale factor that does not depend on the horizon h. Our argument was nonparametric and did not assume the validity of a specific structural model.

2.2 Extension: Non-recursive specifications

Our equivalence result extends straightforwardly to the case of non-recursively identified VARs. Above we restricted attention to recursive identification schemes, as the VAR directly contains a measure of the impulse x_t. In a generic structural VAR identification scheme, the impulse is some – not necessarily recursive – rotation of reduced-form forecasting residuals. Thus, let us continue to consider the VAR (2), but now we shall study the propagation of *some* rotation of the reduced-form forecasting residuals,

$$\bar{\eta}_t \equiv b' u_t. \tag{7}$$

where b is a vector of the same dimension as w_t. Under Assumption 1, we can follow the same steps as in Section 2.1 to establish that the VAR-implied impulse response at horizon h of y_t with respect to the innovation $\bar{\eta}_t$ equals – up to scale – the coefficient $\bar{\beta}_h$ of the linear projection

$$y_{t+h} = \bar{\mu}_h + \bar{\beta}_h(b' w_t) + \sum_{\ell=1}^{\infty} \bar{\delta}'_{h,\ell} w_{t-\ell} + \bar{\xi}_{h,t}, \tag{8}$$

where the coefficients are least-squares projection coefficients and the last term is the projection residual. Thus, any recursive or non-recursive SVAR(∞) identification procedure is

[5] Observe that $u_{x,t} - \tilde{x}_t = E(x_t \mid r_t, \{w_\tau\}_{-\infty<\tau<t}) - E(x_t \mid \{w_\tau\}_{-\infty<\tau<t}) = E(u_{x,t} \mid r_t, \{w_\tau\}_{-\infty<\tau<t}) = E(u_{x,t} \mid u_{r,t}, \{w_\tau\}_{-\infty<\tau<t}) = E(u_{x,t} \mid u_{r,t})$.

equivalent with a local projection (8) on a particular linear combination $b'w_t$ of the variables in the VAR (and their lags). For recursive orderings, the equivalence takes the particularly simple form in Section 2.1.

2.3 Extension: Finite lag length

Whereas our main equivalence result in Section 2.1 relied on infinite lag polynomials, we now prove an equivalence result that holds when only finitely many lags are used. Specifically, when p lags of the data are included in the VAR and as controls in the LP, the impulse response estimands for the two methods agree out to horizon p, but generally not at higher horizons. Importantly, this result is still entirely nonparametric, in the sense that we do not impose that the true DGP is a finite-order VAR.

First, we define the finite-order LP and VAR estimands. We continue to impose the nonparametric Assumption 1. Consider any lag length p and impulse response horizon h.

1. Local projection. The local projection impulse response estimand $\beta_h(p)$ is defined as the coefficient on x_t in a projection as in (1), except that the infinite sum is truncated at lag p. Again, we interpret all coefficients and residuals as resulting from a least-squares linear projection.

2. VAR. Consider a linear projection of the data vector w_t onto p of its lags (and a constant), i.e., the projection (2) except with the infinite sum truncated at lag p. Let $A_\ell(p)$, $\ell = 1, 2, \dots, p$, and $\Sigma_u(p)$ denote the corresponding projection coefficients and residual variance. Define $A(L; p) \equiv I - \sum_{\ell=1}^{p} A_\ell(p)$ and the Cholesky decomposition $\Sigma_u(p) = B(p)B(p)'$. Define also the inverse lag polynomial $\sum_{\ell=0}^{\infty} C_\ell(p)L^\ell = C(L; p) \equiv A(L; p)^{-1}$ consisting of the reduced-form impulse responses implied by $A(L; p)$. Then the VAR impulse respond estimand at horizon h is defined as

$$\theta_h(p) \equiv C_{n_r+2,\bullet,h}(p)B_{\bullet,n_r+1}(p),$$

cf. the definition in Section 2.1 with $p = \infty$.

Note that the VAR(p) model used to define the VAR estimand above is "misspecified", in the sense that the reduced-form residuals from the projection of w_t on its first p lags are not white noise in general.

We now state the equivalence result for finite p. The statement of the result is a simple generalization of Proposition 1, which can be thought of as the case $p = \infty$.

Proposition 2. *Impose Assumption 1. Define* $\tilde{x}_t(\ell) \equiv x_t - E(x_t \mid r_t, \{w_\tau\}_{t-\ell \leq \tau < t})$ *for all* $\ell = 0, 1, 2, \ldots$. *Let the nonnegative integers* h, p *satisfy* $h \leq p$. *If* $\tilde{x}_t(p) = \tilde{x}_t(p-h)$, *then* $\theta_h(p) = \sqrt{E(\tilde{x}_t(p)^2)} \times \beta_h(p)$.

Proof. Please see Appendix A.1. □

Thus, under the conditions of the proposition, the LP and VAR impulse responses agree at all horizons $h \leq p$, although generally not at horizons $h > p$. This finding would not be surprising if the true DGP were assumed to be a finite-order VAR (as in Jordà, 2005, and Kilian & Lütkepohl, 2017, Ch. 12.8), but we allow for completely general covariance stationary DGPs. The reason why the result still goes through is that a VAR(p) obtained through least-squares projections perfectly captures the autocovariances of the data out to lag p (but not further), and these are precisely what determine the LP estimand. Baek & Lee (2019) prove a similar result for the related but distinct setting of single-equation Autoregressive Distributed Lag models with a white noise exogenous regressor.

Proposition 2 assumes $\tilde{x}_t(p) = \tilde{x}_t(p-h)$ to obtain an exact result, but the conclusion is likely to hold qualitatively under more general conditions. If x_t is a direct measure of a "shock" and thus uncorrelated with all past data, then $\tilde{x}_t(\ell) = x_t$ for all $\ell \geq 0$, so the conclusion of the proposition holds exactly. More generally, the LP estimand projects y_{t+h} onto $\tilde{x}_t(p)$ (and controls); thus, the projection depends on the first $p+h$ autocovariances of the data. The estimated VAR(p) generally does not precisely capture the autocovariances of the data at lags $p+1, \ldots, p+h$, and so the LP and VAR potentially project on different objects. However, at short horizons $h \ll p$, it will usually be the case in empirically relevant DGPs that $\tilde{x}_t(p) \approx \tilde{x}_t(p-h)$, since it is typically only the first few lags of the data that is useful for forecasting x_t. In this case, the conclusion of Proposition 2 will hold approximately. We provide an illustration in Section 2.4.

In conclusion, even if we use "too short" a lag length p, the LP and VAR impulse response estimands only disagree at horizons longer than p. This is a comforting fact in applications where the main questions of interest revolve around short-horizon impulse responses.

2.4 Graphical illustration

We finish the section by illustrating graphically the previous theoretical results. We do so in the context of a particular data generating process: the structural macro model of Smets & Wouters (2007). We abstract from sampling uncertainty and throughout assume that the

Illustration: Population equivalence of VAR and LP estimands

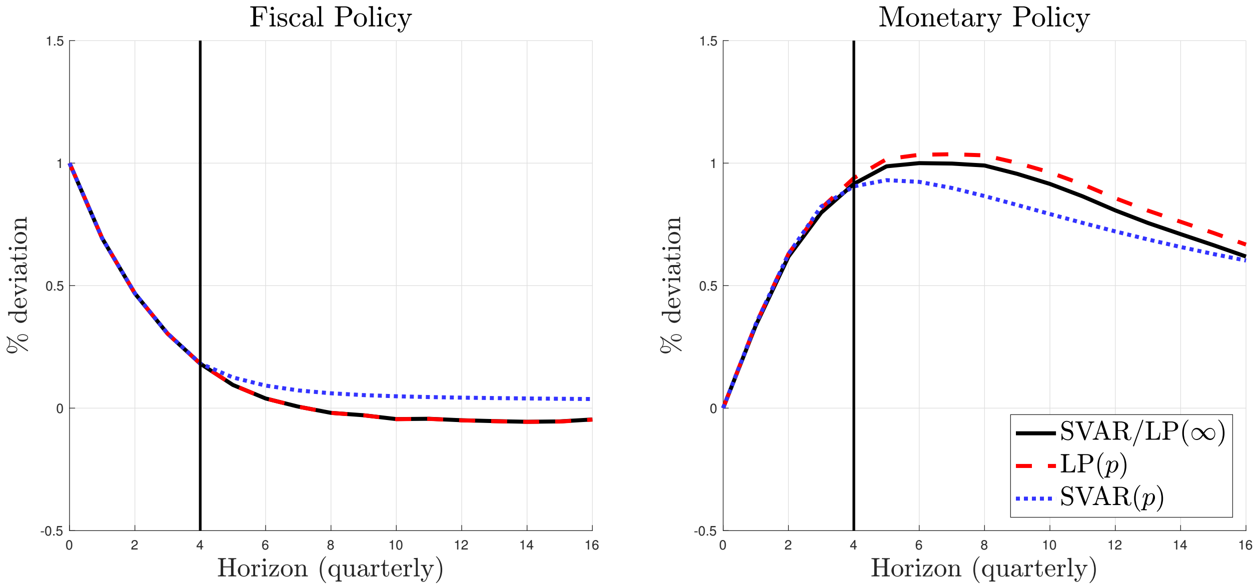

Figure 1: LP and VAR impulse response estimands in the structural model of Smets & Wouters (2007). Left panel: response of output to a government spending innovation. Right panel: response of output to an interest rate innovation. The horizontal line marks the horizon p after which the finite-lag-length LP(p) and VAR(p) estimands diverge.

econometrician actually observes an infinite amount of data.[6] Since this section is merely designed to illustrate the properties of different projections, we do not comment on the relation of the projection estimands to true *structural* model-implied impulse responses. We formally discuss structural identification in Section 4.

The left panel of Figure 1 shows LP and VAR impulse response estimands of the response of output to a government spending innovation. We assume the model's government spending innovation is directly observed by the econometrician, who additionally controls for lags of output and government spending. This experiment is therefore similar in spirit to that of Ramey (2011). As ensured by Proposition 1, the LP(∞) and VAR(∞) estimands – i.e., with infinitely many lags as controls – agree at all horizons. Since by assumption the "impulse" variable x_t is a direct measure of the government spending innovation, we have $\tilde{x}_t(\ell) = x_t$ for all $\ell \geq 0$. Thus, any LP(p) estimand for finite p also agrees with the LP(∞) limit at all horizons. Finally, we observe that the impulse responses implied by a VAR(4) *exactly* agree

[6]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`. To approximate the VAR(∞) representation of the DGP, we truncate the model-implied vector moving average representation at a large horizon ($H = 350$), and then invert.

with the true population projections up until horizon $h = 4$, as predicted by Proposition 2.

The right panel of Figure 1 shows LP and VAR impulse response estimands for the response of output to an innovation in the nominal interest rate. Here the model's innovation is not directly observed by the econometrician, only the interest rate. The LP specifications control for the contemporaneous value of output and inflation as well as lags of output, inflation, and the nominal interest rate; as discussed, this set of control variables is equivalent to ordering the interest rate last in the VAR. Thus, the experiment emulates the familiar monetary policy shock identification analysis of Christiano et al. (2005), although we, at least for the purposes of this section, interpret the projections purely in a reduced-form way. Again, the LP(∞) and VAR(∞) estimands agree at all horizons. Now, however, the "impulse" $\tilde{x}_t(p)$ upon which the different methods project is different. Hence, LP(p) and VAR(p) estimands differ from each other, as well as from the population limit LP(∞)/VAR(∞) estimands. Formally, Proposition 2 only assures that the estimated *impact* impulse responses of LP(p) and VAR(p) agree exactly. Nevertheless, and consistent with the intuition offered in Section 2.3, all impulse response estimands are *nearly* identical until the truncation horizon $p = 4$.

3 Efficient estimation of impulse responses

This section discusses our equivalence result in the context of finite-sample estimation of impulse responses. We first provide a sample analogue of our population equivalence result when the lag length is large. Then we discuss the bias-variance trade-off associated with estimation of impulse response functions.

3.1 Sample equivalence

In addition to being identical *conceptually* and in *population*, we show in Appendix A.2 that local projection and VAR impulse response estimators are nearly identical *in sample* when large lag lengths are used in the regression specifications. Formally, the sample analogue of our result in Section 2 states that the least-squares estimators $\hat{\beta}_h(p)$ and $\hat{\theta}_h(p)$ of the LP and VAR specifications (1)–(2) are likely to be nearly identical (up to scale) at the fixed horizon h, as long as we include a large number p of lags of the data on the right-hand side of the local projection and in the VAR. This result requires certain standard nonparametric regularity conditions, with details also relegated to Appendix A.2.

3.2 Bias-variance trade-off

Empirically relevant short sample sizes force researchers to economize on the number of lags, and the relative accuracy of LP and VAR estimators with a small/moderate number of lags invariably depends on the underlying data generating process (DGP). This is perfectly analogous to the choice between "direct" and "iterated" predictions in multi-step forecasting (Marcellino et al., 2006). Schorfheide (2005) proves that the mean-square error ranking of LP (i.e., direct) and VAR (i.e., iterated) forecasts depends on how large in magnitude the partial autocorrelations of the DGP are at lags longer than the lag length used for estimation.[7] Hence, although Meier (2005), Kilian & Kim (2011), and Choi & Chudik (2019) exhibit simulation evidence that VAR estimators (or other iterated estimators) outperform the LP estimator, this conclusion must necessarily depend on the choice of DGP. Indeed, Brugnolini (2018) and Nakamura & Steinsson (2018) exhibit DGPs where the LP estimator instead outperforms VARs.

More generally, effective finite-sample estimation of impulse responses involves an unavoidable bias-variance trade-off, and many dimension reduction or penalization approaches may be sensible depending on the application. Low-order VARs resolve the bias-variance trade-off by dimension reduction, effectively extrapolating longer-horizon impulse responses from the first few autocorrelations of the data. Bayesian VARs reduce effective dimensionality by imposing priors on longer-lag coefficients, e.g., through a Minnesota prior (Giannone et al., 2015); model averaging across restricted and unrestricted VARs has similar effects (Hansen, 2016). Dimension reduction can also be achieved through penalized local projection (Plagborg-Møller, 2016, Ch. 3; Barnichon & Brownlees, 2018) or by shrinking unrestricted local projections towards low-order VAR estimates (Miranda-Agrippino & Ricco, 2018b). Alternatively, impulse response estimation could be based on plugging a shrinkage/regularized autocovariance function estimate into the explicit formula (3) for the LP/VAR estimand.

We believe that the different estimation methods in the literature are best viewed as sharing the *same* large-sample estimand but lying along a *spectrum* of small-sample bias-variance choices. Low-order VAR(p) models only have a conceptually special status insofar as we think the finite-p assumption is literally true, which is typically not the case. In general, the relative accuracy of the methods depends on smoothness/sparsity properties of the autocovariance function of the data. From the point of view of point estimation, no single method is likely to dominate for *all* empirically relevant data generating processes.

[7] See also Chevillon (2007), McElroy (2015), and references therein.

4 Structural identification

We now argue that our result on the equivalence of LP and VAR impulse response functions has implications for structural identification. We have seen that LP and VAR methods only differ to the extent that they represent different approaches to finite-sample dimensionality reduction. The problem of *structural identification* is a population concept and is thus logically distinct from that of dimensionality reduction. In this section we apply our equivalence result to popular SVAR and local projection identification schemes – including short-run restrictions, long-run restrictions, sign restrictions, and external instruments – and we discuss how to think about non-linear models.

4.1 Structural model

Whereas our previous analysis did not impose any particular structural model, we now impose a linear but otherwise general semiparametric Structural Vector Moving Average (SVMA) model. This model does not restrict the linear transmission mechanism of shocks to observed variables (we address non-linear models in Section 4.4). SVMA models have been analyzed by Stock & Watson (2018), Plagborg-Møller & Wolf (2019), and many others.

Assumption 2. *The data $\{w_t\}$ are driven by an n_ε-dimensional vector $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n_\varepsilon,t})'$ of exogenous structural shocks,*

$$w_t = \mu + \Theta(L)\varepsilon_t, \quad \Theta(L) \equiv \textstyle\sum_{\ell=0}^{\infty} \Theta_\ell L^\ell, \tag{9}$$

where $\mu \in \mathbb{R}^{n_w \times 1}$, $\Theta_\ell \in \mathbb{R}^{n_w \times n_\varepsilon}$, and L is the lag operator. $\{\Theta_\ell\}_\ell$ is assumed to be absolutely summable, and $\Theta(x)$ has full row rank for all complex scalars x on the unit circle. For notational simplicity, we further assume normality of the shocks:

$$\varepsilon_t \overset{i.i.d.}{\sim} N(0, I_{n_\varepsilon}). \tag{10}$$

Under these assumptions w_t is a nonsingular, strictly stationary jointly Gaussian time series, consistent with Assumption 1 in Section 2. The SVMA model is more general than assuming a (stationary) SVAR model or any particular discrete-time linearized DSGE model. The (i, j) element $\Theta_{i,j,\ell}$ of the $n_w \times n_\varepsilon$ moving average coefficient matrix Θ_ℓ is the impulse response of variable i to shock j at horizon ℓ.

The researcher is interested in the propagation of the structural shock $\varepsilon_{1,t}$ to the observed macro aggregate y_t. Since y_t is the (n_r+2)-th element in w_t, the parameters of interest are $\Theta_{n_r+2,1,h}$, $h=0,1,2,\ldots$. We will also consider *relative* impulse responses $\Theta_{n_r+2,1,h}/\Theta_{n_r+1,1,0}$. This may be interpreted as the response in y_{t+h} caused by a shock $\varepsilon_{1,t}$ of a magnitude that raises x_t by one unit on impact. Such relative impulse responses are frequently reported in applied work.

4.2 Structural identification and estimation

In this section we review the class of assumptions guaranteeing correct causal identification, with the goal of illustrating how LP methods are as applicable as VAR methods when implementing common identification schemes. Our main result in Section 2.1 implies that LP-based causal estimation can succeed if and only if SVAR-based estimation can succeed, as we now demonstrate.

Identification under invertibility. Standard SVAR analysis assumes (partial) invertibility – that is, the ability to recover the structural shock of interest, $\varepsilon_{1,t}$, as a function of only current and past macro aggregates:

$$\varepsilon_{1,t} \in \operatorname{span}\left(\{w_\tau\}_{-\infty<\tau\leq t}\right). \tag{11}$$

A given SVAR identification scheme then identifies as the candidate structural shock a particular linear combination of the Wold forecasting errors:

$$\tilde{\varepsilon}_{1,t} \equiv b'u_t, \tag{12}$$

where the chosen identification scheme gives the vector b as a function of the reduced-form VAR parameters $(A(L),\Sigma_u)$, or equivalently the Wold decomposition parameters $(C(L),\Sigma_u)$. Under invertibility, there must exist a vector b such that $\tilde{\varepsilon}_{1,t}=\varepsilon_{1,t}$, so SVAR identification can in principle succeed (Fernández-Villaverde et al., 2007; Wolf, 2018).

We now illustrate that common SVAR identification schemes are equally as simple to implement using LP methods. We first consider a standard recursive scheme covered by our benchmark analysis in Section 2.1, and then two more sophisticated approaches requiring the general equivalence result of Section 2.2.

Example 1 (Recursive identification)**.** Christiano et al. (2005) identify monetary policy shocks through a recursive ordering, with the federal funds rate ordered after output, consumption, investment, wages, productivity, and a price deflator, and before profits and money growth. In the notation of Section 2.1, this ordering corresponds to the federal funds rate as the impulse variable x_t, all aggregates ordered before the federal funds rate as the controls r_t, and all other variables in the vector q_t. Under these definitions, the estimand of the local projection (2) is exactly identical to the SVAR estimand of Christiano et al. (2005). In finite samples, the mean-square error ranking of finite-order VAR estimators and finite-order LP estimators (or other equally plausible dimension reduction techniques) is ambiguous, as discussed in Section 3. □

Example 2 (Long-run identification)**.** Blanchard & Quah (1989) identify the effects of demand and supply shocks using long-run restrictions in a bivariate system. Let gdp_t and unr_t denote log real GDP (in levels) and the unemployment rate, respectively. Then $\Delta gdp_t \equiv gdp_t - gdp_{t-1}$ is log GDP growth. Assume $w_t \equiv (\Delta gdp_t, unr_t)'$ follows the SVMA model in Assumption 2 with $n_\varepsilon = 2$ shocks, where the first shock is a supply shock and the second shock a demand shock. Following Blanchard & Quah, assume that both shocks are invertible, cf. (11), and that the long-run effect of the demand shock on the *level* of output is zero, i.e., $\sum_{\ell=0}^{\infty} \Theta_{1,2,\ell} = 0$. Given a large horizon H, consider the linear projection

$$gdp_{t+H} - gdp_{t-1} = \tilde{\mu}_H + \textstyle\sum_{\ell=0}^{\infty} \tilde{\delta}'_{H,\ell} w_{t-\ell} + \tilde{\xi}_{H,t}. \tag{13}$$

We show in Appendix A.3 that, in the limit as $H \to \infty$, the impulse responses $\Theta_{i,1,\ell}$ with respect to the supply shock can be obtained – up to scale – from the local projection (8) with $b = \tilde{\delta}_{H,0}$ and with y_t given by the response variable of interest (either Δgdp_t or unr_t). The scale factor does not depend on the impulse horizon or on the response variable.[8] Hence, *relative* impulse responses $\Theta_{i,1,h}/\Theta_{1,1,0}$ are correctly identified.[9] In finite samples, the mean-square error performance of the proposed procedure relative to the conventional SVAR(p) approach of Blanchard & Quah (1989) will depend on the tuning parameters H and p, and on whether the low-frequency properties of the data are well approximated by a low-order VAR model.[10] □

[8]The impulse responses with respect to the demand shock are also readily obtained once the supply shock has been identified. Up to scale and sign, the researcher can consider any vector $\tilde{b}$ such that $\tilde{b}'b = 0$, and then implement the local projection (8) with $\tilde{b}$ in lieu of b.

[9]Absolute impulse responses can be identified by rescaling the identified shock so it has variance 1.

[10]Christiano et al. (2006) and Mertens (2012) make the related point that SVAR-based long-run identification need not rely on the VAR-implied long-run variance matrix. Alternative nonparametric estimators of the latter may have attractive bias-variance properties, depending on the true DGP.

Example 3 (Sign identification)**.** Uhlig (2005) set-identifies the effects of monetary policy shocks by sign-restricting impulse responses. Suppose we assume invertibility, and we are interested in the impulse response of y_t to a monetary shock at horizon h. Following the logic in Section 2.2, this impulse response is given by $\nu'\check{\beta}_h$ for some unknown vector $\nu \in \mathbb{R}^{n_w}$, where the coefficient vector $\check{\beta}_h$ is obtained from the least-squares projection

$$y_{t+h} = \check{\mu}_h + \check{\beta}_h' w_t + \sum_{\ell=1}^{\infty} \check{\delta}_{h,\ell}' w_{t-\ell} + \check{\xi}_{h,t}.$$

As a simple example of sign restrictions, suppose we restrict the variable r_t (here assumed to be a scalar) to respond positively to a monetary shock at all horizons $s = 0, 1, \ldots, \bar{H}$. For each horizon $s = 0, 1, \ldots, \bar{H}$, store the coefficient vector $\ddot{\beta}_s$ from the projection

$$r_{t+s} = \ddot{\mu}_s + \ddot{\beta}_s' w_t + \sum_{\ell=1}^{\infty} \ddot{\delta}_{s,\ell}' w_{t-\ell} + \ddot{\xi}_{s,t}.$$

Then the *largest possible* response of y_{t+h} to a monetary shock that raises r_t by one unit on impact can be obtained as the solution to the linear program

$$\sup_{\nu \in \mathbb{R}^{n_w}} \nu'\check{\beta}_h \quad \text{subject to} \quad \begin{aligned} &\ddot{\beta}_0'\nu = 1, \\ &\ddot{\beta}_s'\nu \geq 0, \;\; s = 1, \ldots, \bar{H}. \end{aligned}$$

To compute the *smallest possible* impulse response, replace the supremum with an infimum.[11] It is straight-forward to implement more complicated identification schemes by adding additional equality or inequality constraints of the above type.[12] □

These three examples demonstrate that invertibility-based identification need not be thought of as "SVAR identification", contrary to standard practice in textbooks and parts of the literature. We acknowledge, however, that certain identification schemes may be easier to implement in practice using low-order VARs than using LPs, and *vice versa*.

[11] We focus on computing the *bounds* of the identified set. An alternative approach is to sample from the identified set, as is commonly done in the Bayesian SVAR literature (Rubio-Ramírez et al., 2010).

[12] To consider impulse responses to a *one-standard-deviation* monetary shock, replace the equality constraint in the linear program by the constraint $\nu' \operatorname{Var}(u_t)^{-1}\nu = 1$. The resulting linear-quadratic program with inequality constraints is similar to those in Gafarov et al. (2018) and Giacomini & Kitagawa (2018).

Beyond invertibility. If the invertibility assumption (11) is violated, then identification strategies that erroneously assume invertibility – independent of whether they are implemented using VARs, LPs, or any other dimensionality reduction technique – will not measure the true impulse responses.[13] Instead, these methods will measure the impulse responses to a white noise disturbance that is a linear combination of current and *lagged* true structural shocks:

$$\tilde{\varepsilon}_{1,t} = \vartheta(L)\varepsilon_t. \tag{14}$$

The properties of the lag polynomial $\vartheta(L)$ are characterized in detail in Fernández-Villaverde et al. (2007) and Wolf (2018). Combining (9) and (14), we see that, in general, both LP and VAR impulse response estimands are linear combinations of contemporaneous and lagged true impulse responses. Thus, projection on a given identified impulse $\tilde{\varepsilon}_{1,t}$ correctly identifies impulse response functions (up to scale) if and only if $\tilde{\varepsilon}_{1,t}$ affects the response variable y_t *only* through the contemporaneous true structural shock $\varepsilon_{1,t}$. Trivially, this is the case if $\tilde{\varepsilon}_{1,t}$ is a function only of $\varepsilon_{1,t}$ (the invertible case); less obviously, the same is also true if $\tilde{\varepsilon}_{1,t}$ is only contaminated by shocks that do not directly affect the response variable y_t.[14] Instrumental variable identification, discussed next, is the leading example of this second case.

Summary. Structural identification and estimation can be carried out using either LP or VAR techniques, provided that the correct control variables are used. As a matter of identification (i.e., in population), the two methods succeed or fail together. The performance of different dimension reduction techniques in finite samples, as discussed in Section 3, is logically distinct from structural identification. There is therefore no conceptual or practical reason to limit discussion of structural identification to finite-order SVAR(p) models, as commonly done in the literature.

4.3 Identification and estimation with instruments

Instruments (also known as proxy variables) are popular in semi-structural analysis. We here use our main result in Section 2 to show that the influential Local Projection Instrumental

[13] Several recent papers have demonstrated how to perform valid semi-structural identification without assuming invertibility, cf. the references in Plagborg-Møller & Wolf (2019). Often such methods rely on LP or VAR techniques to compute relevant linear projections, without interpreting the VAR disturbances (i.e., Wold innovations) as linear combinations of the contemporaneous true shocks.

[14] In particular, this means that neither invertibility nor recoverability (as defined in Plagborg-Møller & Wolf, 2019) are *necessary* for successful semi-structural inference on impulse response functions.

Variable estimation procedure is equivalent to estimating a VAR with the instrument ordered first. This is true irrespective of the underlying structural model.

An instrumental variable (IV) is defined as an observed variable z_t that is contemporaneously correlated *only* with the shock of interest $\varepsilon_{1,t}$, but not with other shocks that affect the macro aggregate y_t of interest (Stock, 2008; Stock & Watson, 2012; Mertens & Ravn, 2013).[15] More precisely, given Assumption 2, the IV exclusion restrictions are that $E(z_t \varepsilon_{j,\tau}) \neq 0$ if and only if both $j = 1$ and $t = \tau$. Without loss of generality, we can use projection notation to phrase these restrictions as follows.

Assumption 3.

$$z_t = c_z + \textstyle\sum_{\ell=1}^{\infty}(\Psi_\ell z_{t-\ell} + \Lambda_\ell w_{t-\ell}) + \alpha \varepsilon_{1,t} + v_t, \tag{15}$$

where $\alpha \neq 0$, $c_z, \Psi_\ell \in \mathbb{R}$, $\Lambda_\ell \in \mathbb{R}^{1 \times n_w}$, $v_t \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$, and v_t is independent of ε_t at all leads and lags. The lag polynomial $1 - \sum_{\ell=1}^{\infty} \Psi_\ell L^\ell$ is assumed to have all roots outside the unit circle, and $\{\Lambda_\ell\}_\ell$ is absolutely summable.

Crucially, the assumption allows the IV to be contaminated by the independent measurement error v_t. In some applications, we may know by construction of the IV that the lag coefficients Ψ_ℓ and Λ_ℓ are all zero; obviously, such additional information will not present any difficulties for any of the arguments that follow.

The Local Projection Instrumental Variable (LP-IV) approach estimates the impulse responses to the first shock using a two-stage least squares version of LP. Loosely, Mertens (2015), Jordà et al. (2015, 2018), Leduc & Wilson (2017), Ramey & Zubairy (2018), and Stock & Watson (2018) propose to estimate the LP equation (1) using z_t as an IV for x_t. To describe the two-stage least-squares estimand in detail, define $W_t \equiv (z_t, w_t')'$ and consider the "reduced-form" IV projection

$$y_{t+h} = \mu_{RF,h} + \beta_{RF,h} z_t + \sum_{\ell=1}^{\infty} \delta'_{RF,h,\ell} W_{t-\ell} + \xi_{RF,h,t} \tag{16}$$

for any $h \geq 0$. Consider also the "first-stage" IV projection[16]

$$x_t = \mu_{FS} + \beta_{FS} z_t + \sum_{\ell=1}^{\infty} \delta'_{FS,\ell} W_{t-\ell} + \xi_{FS,t}. \tag{17}$$

[15] We focus on the case of a single IV. If multiple IVs for the same shock are available, Plagborg-Møller & Wolf (2019) show that (i) the model is testable, and (ii) all the identifying power of the IVs is preserved by collapsing them to a certain (single) linear combination.

[16] As always, the coefficients and residuals in (16)–(17) should be interpreted as linear projections.

Notice that the first stage does not depend on the horizon h. As in standard cross-sectional two-stage least-squares estimation, the LP-IV estimand is then given by the ratio $\beta_{LPIV,h} \equiv \beta_{RF,h}/\beta_{FS}$ of reduced-form to first-stage coefficients (e.g. Angrist & Pischke, 2009, p. 122).[17]

Stock & Watson (2018) show that, under Assumptions 2 and 3, the LP-IV estimand $\beta_{LPIV,h}$ correctly identifies the *relative* impulse response $\Theta_{n_r+2,1,h}/\Theta_{n_r+1,1,0}$. Importantly, this holds whether or not the shock of interest $\varepsilon_{1,t}$ is invertible in the sense of equation (11).

We now use our main result from Section 2.1 to show that the LP-IV impulse responses can equivalently be estimated from a recursive VAR that orders the IV first. As in Section 2, this result is nonparametric and assumes nothing about the underlying structural model or about the IV z_t.

Corollary 1. *Let Assumption 1 hold for the expanded data vector $W_t \equiv (z_t, w_t')'$ in place of w_t. Consider a recursively ordered SVAR(∞) in the variables $(z_t, w_t')'$, where the instrument is ordered first (the ordering of the other variables does not matter). Let $\tilde{\theta}_{y,h}$ be the SVAR-implied impulse response at horizon h of y_t with respect to the first shock. Let $\tilde{\theta}_{x,0}$ be the SVAR-implied impact impulse response of x_t with respect to the first shock.*

Then $\tilde{\theta}_{y,h}/\tilde{\theta}_{x,0} = \beta_{LPIV,h}$.

Proof. Let $\tilde{z}_t \equiv \alpha\varepsilon_{1,t} + v_t$ and $a \equiv \sqrt{E(\tilde{z}_t^2)} = \sqrt{\alpha^2 + \sigma_v^2}$. Proposition 1 states that $\tilde{\theta}_{y,h} = a \times \beta_{RF,h}$ for all h, and $\tilde{\theta}_{x,0} = a \times \beta_{FS}$. The claim follows. □

This nonparametric result implies that, given the *structural* Assumptions 2 and 3, valid identification of relative impulse responses can be achieved through *either* LP-IV or through a recursive SVAR with the IV ordered first.[18] Importantly, under Assumptions 2 and 3, these equivalent estimation strategies are valid even when the shock of interest $\varepsilon_{1,t}$ is not invertible (Stock & Watson, 2018). Intuitively, when the IV z_t is added to the VAR, the only reason that the shock $\varepsilon_{1,t}$ may be non-invertible with respect to the *expanded* information set $\{z_\tau, w_\tau\}_{-\infty<\tau\leq t}$ is the presence of the measurement error v_t in the IV equation (15).[19]

[17]In the over-identified case with multiple IVs, the IV estimand can no longer be written as this simple ratio; we focus on a single IV as in most of the applied literature. Moreover, in this subsection we discuss only population equivalence and abstract from finite-sample issues such as weak instruments.

[18]Plagborg-Møller & Wolf (2019) show that point identification of *absolute* impulse responses – and thus variance decompositions – can be achieved under a further *recoverability* assumption that is mathematically and substantively weaker than assuming invertibility.

[19]Note that, even though Assumption 3 allows z_t to be correlated with lags of w_t, non-invertibility of $\varepsilon_{1,t}$ is entirely consistent with Theorem 1 of Stock & Watson (2018). That theorem merely states that if the shock is non-invertible, then it is *possible to construct an example* of an IV $\check{z}_t$ satisfying $E(\check{z}_t\varepsilon_{j,t}) = 0$ for all $j \neq 1$ and $E(\check{z}_t\varepsilon_{j,t-\ell} \mid \{w_\tau\}_{\tau<t}) \neq 0$ for some j and $\ell \geq 1$ (so $\check{z}_t$ does not satisfy Assumption 3).

But this independent measurement error merely leads to attenuation bias in the estimated impulse responses, and the bias (in percentage terms) is the same at all response horizons and for all response variables. Thus, it does not contaminate estimation of *relative* impulse responses. Noh (2018, Proposition 3) derives a result that is similar to our corollary, but his proof method does not exploit the general nonparametric equivalence we establish in Section 2.

IV identification is therefore an example of a setting where SVAR analysis works even though invertibility fails (including the partial invertibility notion of Forni et al., 2018, and Miranda-Agrippino & Ricco, 2018a).[20] Our result implies that it is valid to include an externally identified shock in a SVAR even if the shock is measured with (independent) error, as long as the noisily measured shock is ordered first.[21]

Unlike the non-invertibility-robust procedure of ordering the IV first in a VAR, the popular SVAR-IV procedure of Stock & Watson (2012) and Mertens & Ravn (2013) is only valid under invertibility. This procedure uses an SVAR to identify the shock of interest as

$$\tilde{\varepsilon}_{1,t} \equiv \frac{1}{\sqrt{\mathrm{Var}(\tilde{z}_t^\dagger)}} \times \tilde{z}_t^\dagger,$$

where $\tilde{z}_t^\dagger$ is computed as a linear combination of the reduced-form residuals u_t from a VAR in w_t alone (i.e., excluding the IV from the VAR):

$$\tilde{z}_t^\dagger \equiv E(\tilde{z}_t \mid u_t) = E(\tilde{z}_t \mid \{w_\tau\}_{-\infty<\tau\leq t}).$$

If Assumptions 2 and 3 and the invertibility condition (11) hold, then SVAR-IV is valid. In fact, in this case SVAR-IV removes any attenuation bias, thus correctly identifying *absolute* (not just relative) impulse responses.[22] However, in the general non-invertible case, SVAR-IV mis-identifies the shock as $\tilde{\varepsilon}_{1,t} \neq \varepsilon_{1,t}$.[23] Plagborg-Møller & Wolf (2019, Appendix B.4)

[20] Note that if the invertibility condition (11) fails, then also $\varepsilon_{1,t} \notin \mathrm{span}\left(\{z_\tau, w_\tau\}_{-\infty<\tau\leq t}\right)$ due to the presence of the measurement error v_t in equation (15).

[21] Romer & Romer (2004) and Barakchian & Crowe (2013) include an externally identified monetary shock in a SVAR, but they order it last, which assumes additional exclusion restrictions. Kilian (2006), Ramey (2011), Miranda-Agrippino (2017), and Jarociński & Karadi (2018), among others, mention the strategy of ordering an IV first in a SVAR, but these papers do not consider the non-invertible case.

[22] Consistent with our analytical results, Carriero et al. (2015) observe in a calibrated simulation study that, under invertibility, SVAR-IV correctly identifies *absolute* impulse response functions, while direct projections on the IV suffer from attenuation bias.

[23] The VARX procedure of Paul (2018) has the same estimand as SVAR-IV under his assumptions.

characterize the bias of SVAR-IV under non-invertibility and show that the invertibility assumption can be tested using the IV.

To summarize, the relative impulse responses obtained from the LP-IV procedure of Stock & Watson (2018) are nonparametrically identical to the relative impulse responses from a recursive SVAR with the IV ordered first (an "internal instruments" approach). Assuming an SVMA model and the IV exclusion restrictions, these procedures correctly identify relative structural impulse responses, irrespective of the invertibility of the shock of interest. In contrast, the SVAR-IV procedure of Stock & Watson (2012) and Mertens & Ravn (2013) (an "external instruments" approach) requires invertibility.

4.4 Estimands in non-linear models

Our main result in Section 2.1 implies that linear local projections are exactly as "robust to non-linearities" as VAR methods, in large samples. We now show that the common LP/VAR estimand can be given a mathematically well-defined "best linear approximation" interpretation when the true underlying structural DGP is in fact non-linear.

Assume that the underlying structural DGP has the nonparametric causal structure

$$w_t = g(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \tag{18}$$

where $g(\cdot)$ is any non-linear function that yields a well-defined covariance stationary process $\{w_t\}$, and $\{\varepsilon_t\}$ is an n_ε-dimensional i.i.d. process with $\text{Cov}(\varepsilon_t) = I_{n_\varepsilon}$. The number of structural shocks ε_t may exceed the number of variables in w_t.

We show formally in Appendix A.4 that we can represent the process (18) as the *linear* Structural Vector Moving Average model

$$w_t = \mu^* + \textstyle\sum_{\ell=0}^{\infty} \Theta_\ell^* \varepsilon_{t-\ell} + \sum_{\ell=0}^{\infty} \Psi_\ell^* \zeta_{t-\ell},$$

where ζ_t is an n_w-dimensional white noise process that is uncorrelated at all leads and lags with the structural shocks ε_t. The argument exploits the Wold decomposition of the residual of w_t after projecting on the structural shocks. Hence, the linear SVMA model (9) in Assumption 2 should not be thought of as restrictive, provided we do not restrict the number of "shocks" relative to the number of variables.

The linear SVMA impulse responses Θ_ℓ^* corresponding to the structural shocks ε_t have a "best linear approximation" interpretation. Specifically,

$$(\Theta_0^*, \Theta_1^*, \dots) \in \underset{(\tilde{\Theta}_0, \tilde{\Theta}_1, \dots)}{\operatorname{argmin}} \; E\left[\left(g(\varepsilon_t, \varepsilon_{t-1}, \dots) - \textstyle\sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \varepsilon_{t-\ell}\right)^2\right]. \tag{19}$$

Thus, if a second-moment LP/VAR identification scheme is known to correctly identify the impulse responses in a linear SVMA model (9), and there is doubt about whether the true underlying DGP is in fact linear, the population estimand of the identification procedure can be given a formal "best linear approximation" interpretation. This is analogous to the "best linear predictor" property of Ordinary Least Squares in cross-sectional regression. In contrast, identification approaches that depart from standard linear projections – such as identification through higher moments or through heteroskedasticity – may not have a clear interpretation under functional form misspecification.

We do not take a stand on whether the best linear approximation (19) is of structural interest. In some applications the non-linearities of the true underlying DGP may be of interest *per se*. In such cases, non-linear VAR or LP estimators can be applied, for example by adding interaction or polynomial terms, regime switching, stochastic volatility, etc. Such issues are outside the scope of this paper, which deals exclusively with linear estimators.

5 Empirical application

We finally illustrate our theoretical equivalence results by empirically estimating the dynamic response of corporate bond spreads to a monetary policy shock. We adopt the specification of Gertler & Karadi (2015), who, using high-frequency financial data, obtain an external instrument for monetary policy shocks.[24] Because of possible non-invertibility (Ramey, 2016; Plagborg-Møller & Wolf, 2019), we do not consider the external SVAR-IV estimator, but instead implement direct projections on the IV through (i) local projections and (ii) an "internal instruments" recursive VAR, following the logic of Corollary 1. In both cases, our vector of macro control variables exactly follows Gertler & Karadi (2015); it includes output growth (log growth rate of industrial production), inflation (log growth rate of CPI inflation), the 1-year government bond rate, and the Excess Bond Premium of Gilchrist & Zakrajšek

[24]The external IV z_t is constructed from changes in 3-month-ahead futures prices written on the Federal Funds Rate, where the changes are measured over short time windows around Federal Open Market Committee monetary policy announcement times. See Gertler & Karadi (2015) for details on the construction of the IV and a discussion of the exclusion restriction.

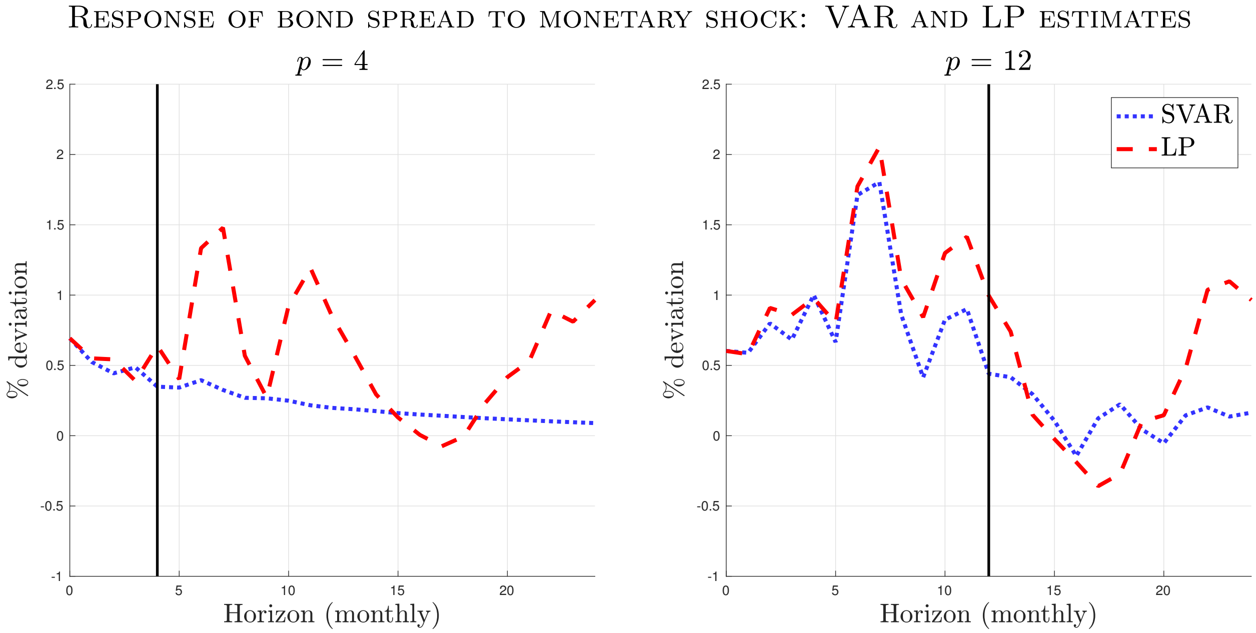

Figure 2: Estimated impulse response function of the Excess Bond Premium to a monetary policy shock, normalized to increase the 1-year bond rate by 100 basis points on impact. Left panel: lag length $p = 4$. Right panel: $p = 12$. The horizontal line marks the horizon p after which the VAR(p) and LP(p) estimates may diverge substantially.

(2012) as a measure of the non-default-related corporate bond spread. The data is monthly and spans January 1990 to June 2012.

Figure 2 shows that LP-IV and "internal instruments" VAR impulse response estimates agree at short horizons, but diverge at longer horizons, consistent with Proposition 2. The figure shows point estimates of the response of the Excess Bond Premium to the monetary policy shock, for different projection techniques and different lag lengths. For all specifications, the Excess Bond Premium initially increases after a contractionary monetary policy shock, consistent with the results in Gertler & Karadi (2015). The left panel shows results for LP(4) and VAR(4) estimates. Up until horizon $h = 4$, the estimated impulse responses are closely aligned. At longer horizons, the iterated VAR structure enforces a smooth return to 0, while direct local projections give more erratic impulse responses. The right panel shows an analogous picture for LP(12) and VAR(12) estimates: The estimated impulse responses agree closely until horizon $h = 12$, but they diverge at longer horizons.

These results provide a concrete empirical illustration of our earlier claim that LP and VAR estimates are closely tied together at short horizons, not just in population but also in sample. The larger the lag length used for estimation, the more impulse response horizons will exhibit agreement between LP and VAR estimates. As this exercise is merely meant to illustrate our theoretical results, we refrain from conducting formal statistical tests of the relative finite-sample efficiency of the different estimation methods.

6 Conclusion

We demonstrated a general *nonparametric* equivalence of local projection and VAR impulse response function estimands. This result has several implications for empirical practice:

1. VAR and local projection estimators of impulse responses should not be regarded as conceptually distinct methods – in population, they estimate the same thing, as long as we control flexibly for lagged data.

2. Efficient finite-sample estimation requires navigating a bias-variance trade-off. Low-order VAR and local projection estimators resolve this trade-off differently, and several other recently proposed methods also lie on the continuum of possible dimension reduction or regularization approaches. Neither low-order VARs nor low-order local projections should be treated as having special status *generally*.

3. The bias-variance trade-off is equivalent to the well-known trade-off between direct and iterated forecasts. Thus, the finite-sample mean-square error ranking of different impulse response estimation methods depends on smoothness/sparsity properties of the autocovariance function of the data. No single method dominates for all empirically relevant data generating processes.

4. At short impulse response horizons, the various estimation methods are likely to approximately agree, but at longer horizons the bias-variance trade-off is unavoidable. A VAR estimator with large lag length will give similar results as a local projection, except at very long horizons.

5. It is a useful diagnostic to check if different estimation methods reach similar conclusions. If estimated impulse responses from VARs and local projections differ substantially at longer horizons, it must mean that the sample partial autocorrelations at long lags are not small. This possibly calls into question the validity of the VAR approximation to the distribution of the data, depending on the noisiness of the estimated impulse responses.

6. Structural *identification* is logically distinct from the dimension reduction choices that must be made for *estimation* purposes. It may be counterproductive to follow standard practice in assuming a finite-order SVAR model whenever the discussion turns to structural identification, as this conflates the population identification analysis and the dimension reduction technique of using a low-order VAR estimator.

7. Any structural estimation method that works for SVARs can be implemented with local projections, and *vice versa*. For example, if a paper already relies on local projections for parts of the analysis, then an additional sign restriction identification exercise, say, can also be implemented in a local projection fashion.

8. If an instrument/proxy for the shock of interest is available, structural impulse responses can be consistently estimated by ordering the instrument first in a recursive VAR (an "internal instruments" approach), even if the shock of interest is noninvertible. In contrast, the popular SVAR-IV estimator (an "external instruments" approach) is only consistent under invertibility.

9. Linear local projections are exactly as "robust to non-linearities" in the underlying data generating process as linear VARs.

This paper has focused entirely on identification and estimation of impulse responses using linear methods. Identification of other objects, such as variance/historical decompositions, is more involved, as shown in Plagborg-Møller & Wolf (2019). It is a promising area of future research to apply and adapt the results in the present paper to *nonlinear* estimation approaches and to questions of *inference* about impulse responses, including in problematic cases such as weak instruments and near-unit roots.

A Appendix

A.1 Equivalence result with finite lag length

We here prove Proposition 2 from Section 2.3. We proceed mostly as in the proof of Proposition 1. As a first step, the Frisch-Waugh theorem implies that

$$\beta_h(p) = \frac{\text{Cov}(y_{t+h}, \tilde{x}_t(p))}{E(\tilde{x}_t(p)^2)}. \tag{20}$$

We now introduce the notation $\text{Cov}^p(\cdot, \cdot)$, which denotes covariances of the data $\{w_t\}$ as implied by the (counterfactual) stationary "fitted" SVAR(p) model

$$A(L; p)w_t = B(p)\bar{\eta}_t, \quad \bar{\eta}_t \sim WN(0, I), \tag{21}$$

i.e., where $\bar{\eta}_t$ is truly white noise (unlike the residuals from the VAR(p) projection on the actual data). For example $\text{Cov}^p(y_t, x_{t-1})$ denotes the covariance of y_t and x_{t-1} that *would* obtain if $w_t = (r_t', x_t, y_t, q_t')'$ were generated by the model (21) with parameters $A(L; p)$ and $B(p)$ obtained from the projection on the *actual* data, as defined in Section 2.3. We similarly define any covariances that involve $\bar{\eta}_t$. Note that stationarity of the VAR model (21) follows from Brockwell & Davis (1991, Remark 2, pp. 424–425).

It follows from the argument in Brockwell & Davis (1991, p. 240) that $\text{Cov}^p(w_t, w_{t-h}) = \text{Cov}(w_t, w_{t-h})$ for all $h \leq p$ (see also Brockwell & Davis, 1991, Remark 2, pp. 424–425 for the multivariate generalization of the key step in the argument). In words, the autocovariances implied by the "fitted" SVAR(p) model (21) agree with the autocovariances of the actual data out to lag p, although generally not after lag p.

Under the counterfactual model (21), we have the moving average representation $w_t = C(L; p)B(p)\bar{\eta}_t$, and thus

$$\theta_h(p) = C_{n_r+2,\bullet,h}(p)B_{\bullet,n_r+1}(p) = \text{Cov}^p(y_{t+h}, \bar{\eta}_{x,t}), \tag{22}$$

where $\bar{\eta}_{x,t}$ is the $(n_r + 1)$-th element of $\bar{\eta}_t$. Since $B(p)$ is lower triangular by definition, it is straight-forward to show from (21) that

$$B_{n_r+1,n_r+1}(p)\bar{\eta}_{x,t} = x_t - E^p(x_t \mid r_t, \{w_\tau\}_{t-p\leq\tau<t}) = x_t - E(x_t \mid r_t, \{w_\tau\}_{t-p\leq\tau<t}) = \tilde{x}_t(p), \tag{23}$$

where $E^p(\cdot \mid \cdot)$ denotes linear projection under the inner product $\text{Cov}^p(\cdot, \cdot)$, the third inequality follows from the above-mentioned equivalence of $\text{Cov}^p(\cdot,\cdot)$ and $\text{Cov}(\cdot,\cdot)$ out to lag p, and the last equality follows by definition. Since $\text{Cov}^p(\bar{\eta}_{x,t}, \bar{\eta}_{x,t}) = 1$, equation (23) implies

$$B_{n_r+1,n_r+1}(p)^2 = \text{Cov}^p(\tilde{x}_t(p), \tilde{x}_t(p)) = E(\tilde{x}_t(p)^2),$$

where the last equality again uses the equivalence of $\text{Cov}^p(\cdot,\cdot)$ and $\text{Cov}(\cdot,\cdot)$ out to lag p. Putting together (22), (23), and the above equation, we have shown that

$$\theta_h(p) = \frac{1}{\sqrt{E(\tilde{x}_t(p)^2)}} \times \text{Cov}^p(y_{t+h}, \tilde{x}_t(p)).$$

Under the stated assumption that $\tilde{x}_t(p) = \tilde{x}_t(p-h)$, the covariance on the right-hand side above depends only on autocovariances of the data w_t at lags $\ell = 0, 1, 2, \dots, p$. Hence, we can again appeal to the equivalence of $\text{Cov}^p(\cdot,\cdot)$ with the covariance function of the actual data, and the expression (20) yields the desired conclusion. $\square$

A.2 In-sample near-equivalence of LP and VAR impulse responses

Complementing the informal discussion in Section 3, here we prove that local projections and recursively identified VARs estimate nearly the same impulse response functions *in sample*, provided the lag lengths used in the specifications are large enough. Assume we observe the data $w_1, w_2, \dots, w_T$ (recall the notation in Section 2.1). For all lag lengths $p \leq T$, define the following:

- Let $\hat{x}_t(p)$ be the residual from a regression of x_t on an intercept, r_t, and $w_{t-1}, \dots, w_{t-p}$.
- Let $\hat{\beta}_h(p)$ denote the OLS estimator of the local projection parameter β_h in the sample version of regression equation (1), where we include p lags of w_t on the right-hand side instead of the infeasible infinite distibuted lag. By the Frisch-Waugh theorem,

$$\hat{\beta}_h(p) = \frac{\sum_{t=p+1}^{T-h} y_{t+h}\hat{x}_t(p)}{\sum_{t=p+1}^{T-h} \hat{x}_t(p)^2}.$$

- Let $\hat{\theta}_h(p)$ denote the horizon-h impulse response of y_t to an innovation in x_t in a Cholesky-identified VAR(p) model (with intercept) estimated by least squares on the data points $t = p+1, p+2, \dots, T$.

In detail, the VAR estimator $\hat{\theta}_h(p)$ is defined as follows. Let $\hat{A}_\ell(p)$ denote the usual least-squares VAR(p) coefficient matrix estimator at lag ℓ, and let $\hat{c}(p)$ denote the corresponding intercept vector estimator. Let $\hat{u}_t(p)$ denote the residual vector. Define the innovation covariance matrix estimator $\hat{\Sigma}(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{u}_t(p)\hat{u}_t(p)'$ and let $\hat{\Sigma}(p) = \hat{B}(p)\hat{B}(p)'$ denote its lower triangular Cholesky decomposition. Define the reduced-form impulse response matrices by $\hat{C}_0(p) = I_{n_w}$ and $\hat{C}_m(p) = \sum_{\ell=1}^{m} \hat{A}_\ell(p)\hat{C}_{m-\ell}(p)$ for $m = 1, \dots, h$. Then $\hat{\theta}_h(p)$ is given by the $(n_r + 2, n_r + 1)$ element of $\hat{C}_h(p)\hat{B}(p)$.

Note that the VAR(p) residuals

$$\hat{u}_t(p) \equiv w_t - \hat{c}(p) - \sum_{\ell=1}^{p} \hat{A}_\ell(p) w_{t-\ell}, \quad t = p+1, p+2, \dots, T,$$

satisfy

$$\sum_{t=p+1}^{T} \hat{u}_t(p) = 0_{n_w \times 1}, \quad \sum_{t=p+1}^{T} \hat{u}_t(p) w_{t-\ell} = 0_{n_w \times n_w}, \;\; \ell = 1, 2, \dots, p. \tag{24}$$

We adopt the convention that $\hat{u}_t(p) \equiv 0$ whenever $t \leq p$.

We are now ready to state the near-equivalence result for LP and VAR impulse response estimators. Let $\|\cdot\|$ denote the Frobenius norm.

Proposition 3. *In the following, the lag length $p = p(T)$ used for estimation is implicitly a function of T. Assume the following:*

i) $\{w_t\}$ is covariance stationary and has a VAR(∞) representation (2), where $\sum_{\ell=1}^{\infty} \|A_\ell\| < \infty$, and the Wold innovations u_t have finite and positive definite covariance matrix Σ. (We do not assume that the innovations are necessarily Gaussian.)

ii) $\|\hat{c}(p) - c\| = o_p(1)$, $\|\hat{A}(p) - A(p)\| = o_p(1)$, and $\|\hat{\Sigma}(p) - \Sigma\| = o_p(1)$, where we have defined $\hat{A}(p) \equiv (\hat{A}_1(p), \dots, \hat{A}_p(p))$ and $A(p) \equiv (A_1, \dots, A_p)$.

Then

$$\hat{\theta}_h(p) = \frac{\frac{1}{T-p}\sum_{t=p+1}^{T-h} y_{t+h}\hat{x}_t(p)}{\left(\frac{1}{T-p}\sum_{t=p+1}^{T} \hat{x}_t(p)^2\right)^{1/2}} + O_p(\hat{R}(p)),$$

where

$$\hat{R}(p) \equiv \frac{\max\{1, \sup_{1\leq t \leq T} \|w_t\|\}^2}{T-p} + \left(\sum_{\ell=p-h+1}^{p} \|\hat{A}_\ell(p)\|^2\right)^{1/2}.$$

Proof. See Appendix A.5. □

Thus, the VAR impulse response estimator $\hat{\theta}_h(p)$ approximately equals the LP impulse response estimator $\hat{\beta}_h(p)$ up to a scale factor that does not depend on the horizon h. The approximation error is of an order $O_p(\hat{R}(p))$ that is likely to be small unless the data is so persistent that the estimated VAR coefficients at the very longest lags are non-negligible.

Assumptions (i) and (ii) of the proposition are easily satisfied under standard nonparametric regularity conditions on the data generating process and a restriction on how quickly the lag length p can grow with T. See for example Lewis & Reinsel (1985) and Gonçalves & Kilian (2007).

A.3 Long-run identification using local projections

Here we show that the LP-based long-run identification approach in Example 2 is valid. Define the Wold innovations $u_t \equiv w_t - E(w_t \mid \{w_\tau\}_{-\infty<\tau<t})$ and Wold decomposition

$$w_t = \chi + C(L)u_t, \quad C(L) \equiv I_2 + \textstyle\sum_{\ell=1}^{\infty} C_\ell L^\ell. \tag{25}$$

Since both structural shocks are assumed to be invertible, there exists a 2×2 matrix B such that $\varepsilon_t = Bu_t$. Comparing (9) and (25), we then have $\Theta(1)B = C(1)$. Let $e_1 \equiv (1, 0)'$. Note that the Blanchard & Quah assumption $e_1'\Theta(1) = (\Theta_{1,1}(1), 0)$ implies

$$e_1'C(1) = e_1'\Theta(1)B = \Theta_{1,1}(1)e_1'B,$$

and therefore

$$e_1'C(1)u_t = \Theta_{1,1}(1) \times e_1'Bu_t = \Theta_{1,1}(1) \times \varepsilon_{1,t}.$$

By the result in Section 2.2, the claim in Example 2 follows if we show that

$$\lim_{H\to\infty} \tilde{\delta}_H' = e_1'C(1). \tag{26}$$

Define $\Sigma_u \equiv \mathrm{Var}(u_t)$. Applying the Frisch-Waugh theorem to the projection (13), and using $w_{1,t} = \Delta gdp_t$, we find

$$\tilde{\delta}_H' = \mathrm{Cov}(gdp_{t+H} - gdp_{t-1}, u_t)\Sigma_u^{-1} = \mathrm{Cov}\left(\sum_{\ell=0}^{H} w_{1,t+\ell}, u_t\right)\Sigma_u^{-1} = \sum_{\ell=0}^{H} \mathrm{Cov}(w_{1,t+\ell}, u_t)\Sigma_u^{-1}. \tag{27}$$

On the other hand, the Wold decomposition (25) implies (recall that u_t is white noise)

$$\sum_{\ell=0}^{\infty} \operatorname{Cov}(w_{t+\ell}, u_t)\Sigma_u^{-1} = \sum_{\ell=0}^{\infty} C_\ell = C(1). \tag{28}$$

Comparing (27) and (28), we get the desired result (26). $\square$

A.4 Best linear approximation under non-linearity

Here we give the technical details behind the "best linear approximation" interpretation of a non-linear model, cf. Section 4.4. Assume the nonparametric model (18), and that $\{w_t\}$ is covariance stationary and purely nondeterministic. Let the linear projection of w_t on the orthonormal basis $(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ be denoted $\sum_{\ell=0}^{\infty} \Theta_\ell^* \varepsilon_{t-\ell}$, with projection residual v_t. Assume v_t is either identically zero or purely non-deterministic. Then it has a Wold decomposition

$$v_t = \mu^* + \textstyle\sum_{\ell=0}^{\infty} \Psi_\ell^* \zeta_{t-\ell},$$

where $\{\zeta_t\}$ is n_w-dimensional white noise with $\operatorname{Cov}(\zeta_t) = I_{n_w}$. Since v_t is a function of $\{\varepsilon_\tau\}_{\tau \le t}$, and $\{\varepsilon_t\}$ is i.i.d., we have $\operatorname{Cov}(\varepsilon_{t+\ell}, v_t) = 0_{n_\varepsilon \times n_w}$ for all $\ell \ge 1$. Moreover, since v_t is a residual from a projection onto $\{\varepsilon_\tau\}_{\tau \le t}$, we also have $\operatorname{Cov}(\varepsilon_{t+\ell}, v_t) = 0_{n_\varepsilon \times n_w}$ for all $\ell \le 0$. By the Wold decomposition theorem, ζ_t lies in the closed linear span of $\{v_\tau\}_{\tau \le t}$, so we must have $\operatorname{Cov}(\varepsilon_{t+\ell}, \zeta_t) = 0_{n_\varepsilon \times n_w}$ for all $\ell \in \mathbb{Z}$. Finally, the best linear approximation property (19) is a standard consequence of linear projection. We have thus verified all claims made in Section 4.4. $\square$

A.5 Proof of Proposition 3

We split the proof into several steps.

STEP 1. We will show that $\sum_{\ell=1}^{p} \|\hat{A}_\ell(p)\| = O_p(1)$. The statement follows from

$$\sum_{\ell=1}^{p} \|\hat{A}_\ell(p)\| \le \sum_{\ell=1}^{p} \|A_\ell\| + \sum_{\ell=1}^{p} \|\hat{A}_\ell(p) - A_\ell\| \le \sum_{\ell=1}^{\infty} \|A_\ell\| + \|\hat{A}(p) - A(p)\|$$

and then exploiting assumptions (i) and (ii).

Step 2. We will show that $\sup_{p+1\le t\le T}\|\hat{u}_t(p)\| = \sup_{1\le t\le T}\|w_t\|\times O_p(1)$. Observe that

$$\sup_{p+1\le t\le T}\|\hat{u}_t(p)\| = \sup_{p+1\le t\le T}\left\|w_t - \sum_{\ell=1}^{p}\hat{A}_\ell(p)w_{t-\ell}\right\| \le \left(\sup_{1\le t\le T}\|w_t\|\right)\left(1+\sum_{\ell=1}^{p}\|\hat{A}_\ell(p)\|\right).$$

Step 1 then gives the desired result.

Step 3. We will show that, for any $m=0,1,\dots,h$,

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-m}(p) = O_p\left(\frac{\sup_t\|w_t\|}{T-p}\right).$$

We have

$$\sum_{t=p+1}^{T}\hat{u}_{t-m}(p) = \sum_{t=p+1}^{T}\hat{u}_t(p) - \sum_{t=T-m+1}^{T}\hat{u}_t(p).$$

The first sum on the right-hand side is exactly zero by the orthogonality conditions (24). The second sum consists of m terms, each of which is $O_p(\sup_t\|w_t\|)$ by Step 2.

Step 4. We will show that, for any $m=1,2,\dots,h$,

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)\hat{u}_{t-m}(p)' = O_p\left(\left(\sum_{\ell=p-h+1}^{p}\|\hat{A}_\ell(p)\|^2\right)^{1/2}\right).$$

$\hat{u}_{t-m}(p)$ is a linear function of $w_{t-m}, w_{t-1-m},\dots,w_{t-p-m}$. By the orthogonality conditions (24), $\hat{u}_t(p)$ is orthogonal to $w_{t-m}, w_{t-1-m},\dots,w_{t-p}$ (and a constant). Thus,

$$\left\|\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)\hat{u}_{t-m}(p)'\right\| = \left\|\frac{1}{T-p}\sum_{t=p+m+1}^{T}\hat{u}_t(p)\sum_{\ell=p-m+1}^{p}w_{t-m-\ell}'\hat{A}_\ell(p)'\right\|$$
$$\le \left(\sum_{\ell=p-m+1}^{p}\|\hat{A}_\ell(p)\|^2\right)^{1/2}\left(\sum_{\ell=p-m+1}^{p}\left\|\frac{1}{T-p}\sum_{t=p+m+1}^{T}\hat{u}_t(p)w_{t-m-\ell}'\right\|^2\right)^{1/2}.$$

Note that $\sum_{\ell=p-m+1}^{p} \|\hat{A}_\ell(p)\|^2 \le \sum_{\ell=p-h+1}^{p} \|\hat{A}_\ell(p)\|^2$ since $m \le h$. Finally,

$$\begin{aligned}
\left\| \frac{1}{T-p} \sum_{t=p+m+1}^{T} \hat{u}_t(p) w_{t-m-\ell}' \right\| &\le \left(\frac{1}{T-p} \sum_{t=p+m+1}^{T} \|\hat{u}_t(p)\|^2 \right)^{1/2} \left(\frac{1}{T-p} \sum_{t=p+m+1}^{T} \|w_{t-m-l}\|^2 \right)^{1/2} \\
&\le \left(\frac{1}{T-p} \sum_{t=p+1}^{T} \|\hat{u}_t(p)\|^2 \right)^{1/2} \left(\frac{1}{T-p} \sum_{t=1}^{T} \|w_t\|^2 \right)^{1/2} \\
&\le \|\hat{\Sigma}(p)\| \left(\frac{1}{T-p} \sum_{t=1}^{T} \|w_t\|^2 \right)^{1/2}.
\end{aligned}$$

The first factor on the right-hand side above is $O_p(1)$ by assumption (ii), while the second factor is $O_p(1)$ since $E\|w_t\|^2 < \infty$.

Step 5. Let $\ell, m \ge 0$ satisfy $m \le h$ and $\ell \le p$. If $m \le \ell$, then

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w_{t-\ell}' = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_t(p) w_{t-(\ell-m)}' + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p} \right), \tag{29}$$

while if $m > \ell$, then

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w_{t-\ell}' = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-(m-\ell)}(p) w_t' + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p} \right), \tag{30}$$

where the $O_p(\cdot)$ terms are uniform in ℓ and m. Claim (30) is proven in the same way as (29), so we only prove the latter. Simply note that

$$\sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w_{t-\ell}' = \sum_{t=p+1}^{T} \hat{u}_t(p) w_{t-(\ell-m)}' - \sum_{t=T-m+1}^{T} \hat{u}_t(p) w_{t-(\ell-m)}',$$

and the second sum consists of m terms, each of which is $O_p(\sup_t \|w_t\|^2)$ by Step 2.

Step 6. We will show that, for any $\ell, m \ge 0$ such that $m \le h$ and $m < \ell \le p$,

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w_{t-\ell}' = O_p\left(\frac{\sup_t \|w_t\|^2}{T-p} \right),$$

where the $O_p(\cdot)$ term is uniform in ℓ and m. By Step 5,

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-m}(p)w_{t-\ell}' = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)w_{t-(\ell-m)}' + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p}\right).$$

Since $1 \leq \ell - m \leq p$, the sum on the right-hand side is precisely zero by the orthogonality conditions (24).

Step 7. Define for all $m = 0, 1, \ldots, h$ the matrix $\hat{H}_m(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} w_t \hat{u}_{t-m}(p)'$. We will show that

$$\hat{H}_m(p) = \sum_{\ell=1}^{m}\hat{A}_\ell(p)\hat{H}_{m-\ell}(p) + O_p(\hat{R}(p)), \quad m = 1, 2, \ldots, h.$$

Let $m = 1, \ldots, h$ be arbitrary. Since

$$w_t = \hat{c}(p) + \sum_{\ell=1}^{p}\hat{A}_\ell(p)w_{t-\ell} + \hat{u}_t(p),$$

we obtain

$$\begin{aligned}
\hat{H}_m(p) &= \frac{1}{T-p}\sum_{t=p+1}^{T} w_t\hat{u}_{t-m}(p)' \\
&= \sum_{\ell=1}^{p}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' \\
&\quad + \hat{c}(p)\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-m}(p)' \\
&\quad + \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)\hat{u}_{t-m}(p)'.
\end{aligned}$$

By Step 3, the second term above is $O_p(\frac{1}{T-p}\sup_t \|w_t\|)$. By Step 4, the third term is $O_p((\sum_{\ell=p-h+1}^{p}\|\hat{A}_\ell(p)\|^2)^{1/2})$. As for the first term above, we split it up as follows:

$$\begin{aligned}
\sum_{\ell=1}^{p}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' &= \sum_{\ell=1}^{m}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' \\
&\quad + \sum_{\ell=m+1}^{p}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)'.
\end{aligned}$$

By Steps 1 and 6, the second term above is $O_p(\frac{1}{T-p}\sup_t \|w_t\|^2)$. By Steps 1 and 5, the first term above equals

$$\begin{aligned}\sum_{\ell=1}^{m} \hat{A}_\ell(p) \frac{1}{T-p} \sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' &= \sum_{\ell=1}^{m} \hat{A}_\ell(p) \frac{1}{T-p} \sum_{t=p+1}^{T} w_t \hat{u}_{t-(m-\ell)}(p)' + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p}\right) \\ &= \sum_{\ell=1}^{m} \hat{A}_\ell(p)\hat{H}_{m-\ell}(p) + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p}\right).\end{aligned}$$

Step 8. We will show that $\hat{H}_m(p) = \hat{C}_m(p)\hat{\Sigma}(p) + O_p(\hat{R}(p))$ for all $m = 0, \dots, h$. We proceed by induction on m. The claim is true by definition for $m = 0$. Assume the claim is true for all $m \leq \tilde{m} - 1$. Then Step 7 implies

$$\begin{aligned}\hat{H}_{\tilde{m}} &= \sum_{\ell=1}^{\tilde{m}} \hat{A}_\ell(p)\hat{H}_{\tilde{m}-\ell}(p) + O_p(\hat{R}(p)) \\ &= \sum_{\ell=1}^{\tilde{m}} \hat{A}_\ell(p)\{\hat{C}_{\tilde{m}-\ell}(p)\hat{\Sigma}(p) + O_p(\hat{R}(p))\} + O_p(\hat{R}(p)) \\ &= \left(\sum_{\ell=1}^{\tilde{m}} \hat{A}_\ell(p)\hat{C}_{\tilde{m}-\ell}(p)\right)\hat{\Sigma}(p) + O_p(\hat{R}(p)) \\ &= \hat{C}_{\tilde{m}}(p)\hat{\Sigma}(p) + O_p(\hat{R}(p)).\end{aligned}$$

Here the penultimate equality uses Step 1, and the last equality uses the recursive definition of $\hat{C}_{\tilde{m}}(p)$.

Step 9. We will show that $\|\hat{B}(p)^{-1}\| = O_p(1)$. This follows from assumption (ii), the continuity of the Cholesky decomposition at any positive definite matrix, and the assumption (i) that Σ is positive definite.

Step 10. Let e_x be the $(n_r + 2)$-th n_w-dimensional unit vector, i.e., $x_t = e_x' w_t$. Then

$$e_x'\hat{B}(p)^{-1}\hat{u}_t(p) = \frac{1}{(\frac{1}{T-p}\sum_{t=p+1}^{T} \hat{x}_t(p)^2)^{1/2}}\hat{x}_t(p)$$

for all $t = p+1, p+2, \dots, T$. This is just the sample analogue of the population result (5)–(6), so we refrain from giving the details of the proof.

Step 11. We will show that

$$\hat{C}_h(p)\hat{B}(p)e_x = \frac{1}{(\frac{1}{T-p}\sum_{t=p+1}^T \hat{x}_t(p)^2)^{1/2}} \times \frac{1}{T-p}\sum_{t=p+1}^T w_t \hat{x}_{t-h}(p)' + O_p(\hat{R}(p)).$$

By Steps 8 and 9,

$$\hat{C}_h(p)\hat{B}(p) = \hat{C}_h(p)\hat{\Sigma}(p)\hat{B}(p)^{-1\prime} = \hat{H}_m\hat{B}(p)^{-1\prime} + O_p(\hat{R}_p).$$

Hence,

$$\hat{C}_h(p)\hat{B}(p)e_x = \frac{1}{T-p}\sum_{t=p+1}^T w_t \left(e_x'\hat{B}(p)^{-1}\hat{u}_{t-h}(p)\right)' + O_p(\hat{R}_p),$$

so the claim follows from Step 10.

Step 12. The statement of the proposition follows from Step 11 and the fact that $\hat{\theta}_h(p)$ by definition equals the (n_r+2)-th element of $\hat{C}_h(p)\hat{B}(p)e_x$. □

References

Angrist, J. & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Baek, C. & Lee, B. (2019). ABCs of Understanding Single Equation Regressions. Manuscript, University of California, Berkeley.

Barakchian, S. M. & Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics*, *60*(8), 950–966.

Barnichon, R. & Brownlees, C. (2018). Impulse Response Estimation By Smooth Local Projections. *Review of Economics and Statistics*. Forthcoming.

Blanchard, O. & Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, *79*(4), 655–673.

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

Brugnolini, L. (2018). About Local Projection Impulse Response Function Reliability. Manuscript, University of Rome "Tor Vergata".

Carriero, A., Mumtaz, H., Theodoridis, K., & Theophilopoulou, A. (2015). The Impact of Uncertainty Shocks under Measurement Error: A Proxy SVAR Approach. *Journal of Money, Credit and Banking*, *47*(6), 1223–1238.

Chevillon, G. (2007). Direct Multi-Step Estimation and Forecasting. *Journal of Economic Surveys*, *21*(4), 746–785.

Choi, C.-Y. & Chudik, A. (2019). Estimating impulse response functions when the shock series is observed. *Economics Letters*, *180*, 71–75.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, *113*(1), 1–45.

Christiano, L. J., Eichenbaum, M., & Vigfusson, R. (2006). Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions. *Journal of the European Economic Association*, *4*(2–3), 475–483.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., & Watson, M. W. (2007). ABCs (and Ds) of Understanding VARs. *American Economic Review*, *97*(3), 1021–1026.

Forni, M., Gambetti, L., & Sala, L. (2018). Structural VARs and Non-invertible Macroeconomic Models. *Journal of Applied Econometrics*. Forthcoming.

Frisch, R. (1933). Propagation Problems and Impulse Problems in Dynamic Economics. In *Economic Essays in Honor of Gustav Cassel*. George Allen & Unwin Ltd.

Gafarov, B., Meier, M., & Montiel Olea, J. L. (2018). Delta-Method Inference for a Class of Set-Identified SVARs. *Journal of Econometrics*, *203*(2), 316–327.

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. (2018). Robust Bayesian Inference for Set-Identified Models. Manuscript, University College London.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior Selection for Vector Autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451.

Gilchrist, S. & Zakrajšek, E. (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, *102*(4), 1692–1720.

Gonçalves, S. & Kilian, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Reviews*, *26*(6), 609–641.

Hansen, B. E. (2016). Stein Combination Shrinkage for Vector Autoregressions. Manuscript, University of Wisconsin-Madison.

Jarociński, M. & Karadi, P. (2018). Deconstructing monetary policy surprises: the role of information shocks. ECB Working Paper No. 2133.

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. (2015). Betting the house. *Journal of International Economics*, *96*(Supplement 1), S2–S18. 37th Annual NBER International Seminar on Macroeconomics.

Jordà, Ò., Schularick, M., & Taylor, A. M. (2018). The effects of quasi-random monetary experiments. Manuscript, Federal Reserve Bank of San Francisco.

Kilian, L. (2006). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. CEPR Discussion Paper 5994.

Kilian, L. & Kim, Y. J. (2011). How Reliable Are Local Projection Estimators of Impulse Responses? *Review of Economics and Statistics*, *93*(4), 1460–1466.

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

Leduc, S. & Wilson, D. (2017). Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act. *American Economic Journal: Economic Policy*, *9*(2), 253–292.

Lewis, R. & Reinsel, G. C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal of Multivariate Analysis*, *16*(3), 393–411.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499–526.

McElroy, T. (2015). When are Direct Multi-step and Iterative Forecasts Identical? *Journal of Forecasting*, *34*(4), 315–336.

Meier, A. (2005). How Big is the Bias in Estimated Impulse Responses? A Horse Race between VAR and Local Projection Methods. Manuscript, European University Institute.

Mertens, E. (2012). Are spectral estimators useful for long-run restrictions in SVARs? *Journal of Economic Dynamics and Control*, *36*(12), 1831–1844.

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.

Miranda-Agrippino, S. (2017). Unsurprising Shocks: Information, Premia, and the Monetary Transmission. Manuscript, Bank of England.

Miranda-Agrippino, S. & Ricco, G. (2018a). Identification with External Instruments in Structural VARs under Partial Invertibility. Manuscript, Bank of England.

Miranda-Agrippino, S. & Ricco, G. (2018b). The Transmission of Monetary Policy Shocks. Manuscript, Bank of England.

Nakamura, E. & Steinsson, J. (2018). Appendix to: "Identification in Macroeconomics". *Journal of Economic Perspectives*, *32*(3), 59–86.

Noh, E. (2018). Impulse-response analysis with proxy variables. Manuscript, University of California San Diego.

Paul, P. (2018). The Time-Varying Effect of Monetary Policy on Asset Prices. Federal Reserve Bank of San Francisco Working Paper 2017-09.

Plagborg-Møller, M. (2016). *Essays in Macroeconometrics*. PhD thesis, Harvard University.

Plagborg-Møller, M. & Wolf, C. K. (2019). Instrumental Variable Identification of Dynamic Variance Decompositions. Manuscript, Princeton University.

Ramey, V. A. (2011). Identifying Government Spending Shocks: It's All in the Timing. *Quarterly Journal of Economics*, *126*(1), 1–50.

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.

Ramey, V. A. & Zubairy, S. (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. *Journal of Political Economy*, *126*(2), 850–901.

Romer, C. D. & Romer, D. H. (2004). A New Measure of Monetary Shocks: Derivation and Implications. *American Economic Review*, *94*(4), 1055–1084.

Rubio-Ramírez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, *77*(2), 665–696.

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, *128*(1), 99–136.

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

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. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal*, *128*(610), 917–948.

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

Wolf, C. K. (2018). SVAR (Mis-)Identification and the Real Effects of Monetary Policy. Manuscript, Princeton University.