Comment: Local Projections or VARs? A Primer for Macroeconomists

(by Montiel-Olea, Plagborg-Møller, Quian, and Wolf) [⋆]

Òscar Jordà [†]

25th May 2025

Impulse response analysis permeates much empirical macroeconomics research. In 2024 alone, over 5,000 papers listed in `scholar.google.com` contain the phrase "impulse responses" along with the word "economics.[1]" Sims (1980) originally proposed to fit vector autoregressions (VARs) to data and then to derive the implied impulse response representation. Later, Jordà (2005) proposed using local projections (LPs) to directly estimate impulse responses instead. In the contribution to this volume, Montiel Olea, Plagborg-Møller, Qian, and Wolf (2025) clarify the pros and cons of each approach. They offer practical advice for practitioners in an accessible, yet rigorous manner.

I pretty much agree with nearly all the recommendations and will therefore focus my discussion on four broader themes: (1) the two contrasting ways to think about impulse responses and their meaning; (2) whether we should be reporting *cumulative response ratios* instead of impulse responses; (3) the implications of choosing the lag length in a VAR versus an LP; and (4) the variance-bias trade-off between VARs and LPs. However, it is important to be upfront about some main topics left behind, largely a deeper dive into identification, and more broadly nonlinearities and applications with panel data. These are meaty topics on their own right, with a large literature dedicated to each.[2]

1. What are impulse responses for?

Although Sims (1980) is rightly credited with popularizing impulse response analysis using VARs, economists have long viewed economies as dynamic systems. For example, Frisch (1933) writes,

---

⋆The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank San Francisco, or the Federal Reserve System.

†Federal Reserve Bank of San Francisco; and Department of Economics, University of California, Davis; and CEPR (oscar.jorda@sf.frb.org; ojorda@ucdavis.edu).

[1] So as to eliminate mentions of impulse responses in other sciences.

[2] A brief overview on these topics can be found in Jordà and Taylor (2025).

"The majority of the economic oscillations which we encounter seem to be explained most plausibly as free oscillations. In many cases they seem to be produced by the fact that certain exterior impulses hit the economic mechanism and thereby initiate more or less regular oscillations." This language is clearly reminiscent of the literature on signal processing and understanding this background clarifies the logic of what Sims (1980) had in mind.

Sims's work appeared at a time when real business cycle theory (see, e.g. Kydland and Prescott, 1982; Long Jr and Plosser, 1983) and the rational expectations revolution (Lucas, 1976) were taking hold, planting the seeds for a new generation of models. These models characterized the economic environment where agents made decisions by taking into account, not just the present but also the future, while recognizing that the future is uncertain, and that their actions would affect those of others. This is the essence of dynamic, stochastic, general equilibrium (DSGE) models, arguably the prevalent paradigm in macroeconomics, and whose particular structure will have a critical influence on how we estimate VARs, as I discuss below.

A natural approach to evaluate such models presented itself: "The best descriptive device appears to be the analysis of the system's response to typical random shocks." (Sims, 1980, p. 21). In other words, impulse responses can be used to trace the propagation of a shock over time generated by the model. Moreover, to the extent that the approximate linearized solutions of DSGE models came in the form of systems of stochastic difference (or in continuous time, differential) equations, further cemented the instinctive approach of using VARs to summarize the stochastic properties of the data.

Sims, of course, recognized two limitations of this approach: (1) "The common econometric practice of summarizing distributed lag relations in terms of their implied long run equilibrium behavior is quite misleading in these systems" (Sims, 1980, p.20); and (2) that the residuals across equations where correlated, which hindered causal interpretation of a VAR's estimates and its impulse responses. Initially this lead to the practice of "triangularization," an early example of which consists of proposing a Cholesky ordering for the variables in the system. Since then, several identification approaches have been proposed but the core idea remains: VARs are primarily estimated to compare the dynamic properties of macroeconomic models against those implied by the data. Since models have relatively restrictive dynamics compared to the richer dynamics of VARs, the exercise is best understood as a model-validation approach, rather than a policy evaluation exercise.

In contrast, LPs where initially conceived as a way to characterize impulse responses without resorting to a reference model.[3] In fact, to understand the logic behind LPs, it is helpful to rely on another development that was taking hold around the same time as Sims (1980): the introduction of the *potential outcomes* paradigm (Rubin, 1974, 1977) and the *program evaluation revolution* that followed. The main thrust of this agenda is best encapsulated by (Angrist, 2001): "...causal relationships answer counterfactual questions and are therefore more likely to be of value predicting the effects of

[3]The original title of Jordà (2005) was "Model-free impulse responses."

changing policies or changing circumstances or understanding the past."

A simple way to motivate the main ideas is as follows. Using capital letters for random variables and lowercase letters for their realizations, consider a binary policy variable, $S_t \in \{0, 1\}$, assigned at random (these two assumptions are made to simplify the exposition). We are interested in evaluating the effect of this policy intervention on some outcome, say $Y_t \in \mathbb{R}$. Denote Y_{1t} as the *potential* outcome when $S_t = 1$, and similarly, Y_{0t} when $S_t = 0$. That is, the data that we observe, $\{y_t, s_t\}_{t=1}^T$, come from a latent mixture given by $Y_t = Y_{1t}S_t + Y_{0t}(1 - S_t) = Y_{0t} + (Y_{1t} - Y_{0t})S_t$, where Y_{1t} and Y_{0t} are not directly observable random variables.

A natural calculation involves comparing the means of the outcome under the following two scenarios: when there is an intervention ($S_t = 1$), versus when there isn't ($S_t = 0$), more specifically:

$$\begin{aligned} E[Y_t|S_t = 1] - E[Y_t|S_t = 0] &= E[Y_{1t}|S_t = 1] - E[Y_{0t}|S_t = 0] \\ &= \underbrace{E[Y_{1t}|S_t = 1] - E[Y_{0t}|S_t = 1]}_{\text{average treatment on the treated}} + \underbrace{\{E[Y_{0t}|S_t = 1] - E[Y_{0t}|S_t = 0]\}}_{\text{bias}}, \end{aligned} \quad (1)$$

Consider the second line of this expression. The main difficulty in measuring the *average treatment effect on the treated* or ATT (as the first term of the decomposition is known) stems from the fact that the counterfactual average response given by $E[Y_{0t}|S_t = 1]$ is not directly observable since $Y_t = Y_{1t}$ whenever $S_t = 1$ (and similarly when $S_t = 0$). When S_t is randomly assigned, $E[Y_{0t}|S_t = 1] = E[Y_{0t}|S_t = 0]$ and the bias term is therefore eliminated but in practice, identification will have to be achieved in some manner, whose discussion I leave for another time since Montiel Olea et al. (2025) also do not discuss identification in much detail either.

Instead of Y_t, we could think of comparing Y_{t+h} so that the ATT in my example could be simply calculated by taking the difference in the sample averages of $\{y_{t+h}\}_{t=1}^{T-h}$ for $h = 0, 1, \ldots, H$ depending on whether $s_t = 0$ or 1, say $\overline{y}_1(h) - \overline{y}_0(h)$. And of course, this difference could be articulated as a regression:

$$y_{t+h} = \alpha_h + \beta_h s_t + \epsilon_{t+h}, \quad \text{for} \quad h = 0, 1, \ldots, H \quad (2)$$

where $\hat{\alpha}_h = \overline{y}_0(h)$ and $\hat{\beta}_h = \overline{y}_1(h) - \overline{y}_0(h)$. Once written in regression form, three natural extensions immediately come to mind. One is relaxing the assumption that S_t is binary. Another is the inclusion of exogenous and predetermined variables, such as lags of the outcome, the intervention, and other variables, which I collect in the vector $\boldsymbol{X}_t$. Under the assumption that S_t is randomly assigned, the inclusion of these variables does not affect the estimate of the ATT but makes it more efficient. In more general settings, they play an role in dampening selection concerns and obtaining correct inference. The final extension, which I will not explore here, is that the regression could be estimated using instrumental variable methods to allow for a causal interpretation, for example.

This discussion suggest extending Equation 2 as:

$$y_{t+h} = \alpha_h + \beta_h s_t + \gamma_h \boldsymbol{x}_t + \epsilon_{t+h}. \tag{3}$$

This is just a linear LP, as originally proposed in Jordà (2005). It suggests that an interesting statistic of interest is the answer to the question: What is the counterfactual average effect of a policy intervention on an outcome for the subpopulation of treated observations? Thus, when thinking about the question in this manner, the natural generalization of Equation 3 is to think more generally on ways to characterize $E[Y_{i,t+h}|S_t = j, \boldsymbol{X}_t]$ for $i, j \in \{0, 1\}$. I will provide a simple example below.

For now, the fact that the particular specification of the LP in Equation 3 coincides with the impulse response from a VAR is, to some extent, a happy (though not unexpected) "coincidence." This coincidence facilitates deriving many useful statistical properties for this important case, but it should not be understood to constrain how one should think about LPs and the questions the researcher wants to answer. Rather, it clarifies the connection between the time series tradition and the ideas from the program evaluation literature.

In fact, once one realizes that connection, it becomes clear that the goal of the analyst should be to characterize the *conditional mean functions* in Equation 1. In particular, Cloyne, Jordà, and Taylor (2023) make the simplest assumption possible and specify a linear conditional mean function for the potential outcomes, specifically, $E[Y_{i,t+h}|S_t = j, \boldsymbol{X}_t] = \alpha_h^i + \gamma_h^i E[\boldsymbol{X}_t|S_t = j]$ for $i, j \in \{0, 1\}$. It is clear that $Y_{1,t+h}$ and $Y_{0,t+h}$ are not directly observable so the analyst will typically write down the difference in conditional means $E[Y_{t+h}|S_t = 1, \boldsymbol{X}_t] - E[Y_{t+h}|S_t = 0, , \boldsymbol{X}_t]$ which from Equation 1 we know delivers an estimate of the ATT under random assignment. However, given our assumptions, note that:

$$\begin{aligned} E[Y_{t+h}|S_t = 1, \boldsymbol{X}_t] - E[Y_{t+h}|S_t = 0, , \boldsymbol{X}_t] &= \\ E[Y_{1,t+h}|S_t = 1, \boldsymbol{X}_t] - E[Y_{0,t+h}|S_t = 0, , \boldsymbol{X}_t]. \end{aligned}$$

Adding and subtracting the counterfactual $E[Y_{0,t+h}|S_t = 1, , \boldsymbol{X}_t]$ this expression can be written as:

$$\begin{aligned} E[Y_{1,t+h}|S_t = 1, \boldsymbol{X}_t] - E[Y_{0,t+h}|S_t = 1, , \boldsymbol{X}_t] + \\ E[Y_{0,t+h}|S_t = 1, \boldsymbol{X}_t] - E[Y_{0,t+h}|S_t = 0, , \boldsymbol{X}_t]. \end{aligned}$$

Further, since $E[Y_{i,t+h}|S_t = j, \boldsymbol{X}_t] = \alpha_h^i + \gamma_h^i E[\boldsymbol{X}_t|S_t = j]$ for $i, j \in \{0, 1\}$, it is easy to see that the previous expression becomes:

$$\alpha_h^1 + \gamma_h^1 E[\boldsymbol{X}_t|S_t = 1] - (\alpha_h^0 + \gamma_h^0 E[\boldsymbol{X}_t|S_t = 1]) +$$
$$\alpha_h^0 + \gamma_h^0 E[\boldsymbol{X}_t|S_t = 1] - (\alpha_h^0 + \gamma_h^0 E[\boldsymbol{X}_t|S_t = 0]) =$$

$$\underbrace{(\alpha_h^1 - \alpha_h^0)}_{\beta_h} + \underbrace{(\gamma_h^1 - \gamma_h^0)}_{\boldsymbol{\theta}_h} E[\boldsymbol{X}_t|S_t = 1] + \underbrace{\gamma_h^0 (E[\boldsymbol{X}_t|S_t = 1] - E[\boldsymbol{X}_t|S_t = 0])}_{\text{balance}}. \tag{4}$$

The term β_h becomes the ATT of interest. A test of the null $H_0 : \gamma_h^1 = \gamma_h^0$, or equivalently $H_0 : \boldsymbol{\theta}_h = 0$ becomes a test of *symmetry*, that is, a test that the controls affect the outcome in the same way regardless of whether an intervention takes place. A test of the null $H_0 : E[\boldsymbol{X}_t|S_t = 1] = E[\boldsymbol{X}_t|S_t = 0]$ becomes a test of *balance*, that is, a check for potential selection issues that would induce bias in the causal interpretation of β_h. Whenever there is symmetry and balance, the ATT can be obtained with a usual LP of Equation 3, but otherwise, this will not be the case.

In practice, Equation 4 can be simply estimated by extending the LP in Equation 3 with an additional interaction term:

$$y_{t+h} = \alpha_h + \beta_h s_t + \gamma_h \boldsymbol{x}_t + \boldsymbol{\theta}_h s_t \boldsymbol{x}_t + \epsilon_{t+h}, \tag{5}$$

with $\alpha_h = \alpha_h^0$, $\beta_h = (\alpha_h^1 - \alpha_h^0)$, $\gamma_h = \gamma_h^0$, and $\boldsymbol{\theta}_h = (\gamma_h^1 - \gamma_h^0)$. Several results are worth remarking. First, note that Equation 5 is linear in parameters and thus can be estimated using least-squares or instrumental variable (IV) methods. Second, tests of symmetry and balance can be easily performed using standard asymptotic approximations. Third, even though we have assumed that the conditional mean function is linear, the impulse response is no longer linear as long as $\boldsymbol{\theta}_h \neq 0$. For any value $\boldsymbol{X}_t = \boldsymbol{x}$, the impulse response is $\mathcal{R}_{s \to y}(h) = \beta_h + \boldsymbol{\theta}_h \boldsymbol{x}$. Thus, depending on the value of $\boldsymbol{x}$, it will attenuate or accentuate the average response measured by β_h.

This is just an example of how quickly the departure between LPs and VARs takes place since, to my knowledge, there is no obvious equivalent way to specify a VAR that would deliver the counterfactual interpretation of LPs in general. Moreover, note that nothing restricts the researcher from specifying a more general conditional mean function, i.e., $E[Y_{i,t+h}|S_t = j, \boldsymbol{X}_t] = m(Y_{i,t+h}, S_t, \boldsymbol{X}_t; \phi)$ for some function $m(.)$ only limited by the practitioner's imagination and needs. Here though, it is useful to cite the recent work of Kolesár and Plagborg-Møller (2025), which shows conditions under which linear methods can nevertheless deliver appropriate estimates of an average ATT.

Relative to traditional cross-sectional results, there are several details to be ironed out since the data are a time series. For example, Rambachan and Shephard (2019) rely on an assumption of *no anticipation* or *non-interference* (Cox, 1958; Rubin, 1980). This assumption basically states that potential outcomes do not depend on future treatment paths. And we usually want to conduct the analysis by conditioning on past information. A detailed list of these and other assumptions can be found in Bojinov and Shephard (2019); Rambachan and Shephard (2019).

I conclude this section by making clear that the previous discussion is not meant to be an indictment of VARs. Rather, it is meant to highlight that each approach, VARs and LPs, is meant to address a different type of question. Moreover, the extent to which we are willing to entertain a data generating process (or DGP) as we do with a VAR, it facilitates the derivation of its statistical properties. Unlike VARs however, LPs do not propose a time series model meant to approximate the spectral density of the underlying time series process. They are not generative models but rather a polyvalent strategy to estimate key statistics of interest. A big swath of the literature, including to a lesser extent Montiel Olea et al. (2025), juxtaposes VARs and LPs as if in direct conflict, but I think this is unfortunate as both methods are useful depending on the application. The problems in economics are too complex and too important to be leaving behind any useful tool.

2. Should we be reporting impulse responses anyway?

Unlike many applied microeconomic studies where interventions or treatments are often (though not exclusively) one-offs, macroeconomics data is quite different. For example, a monetary shock generally results, not just in a shift of interest rates on impact, but also in their subsequent evolution. The implication being that the impulse response reflects how the outcome variable responds to the shock on impact as well as the subsequent trajectory of interest rates. Or consider the fiscal multiplier. Ideally we want to measure total dollars gained relative to total dollars spent whenever the government decides to increase spending. A simple example helps illustrate these observations.

Suppose that data are generated from the following model:

$$\begin{cases} S_t &= \rho S_{t-1} + \epsilon_{s,t} \\ Y_t &= \beta S_t + \psi Y_{t-1} + \epsilon_{y,t} \end{cases}, \tag{6}$$

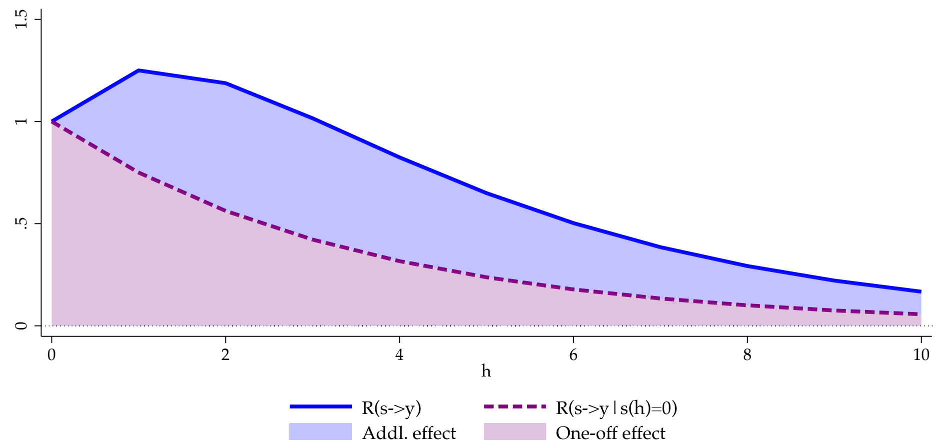
where as before, S_t is an intervention and Y_t is the outcome. Note that S_t is an exogenously determined, though autocorrelated treatment variable. Y_t responds to S_t with a parameter β but it also has internal propagation dynamics given by ψ. Though simple, this setup encompasses a wider class of settings once one views the process for Y_t as representing the companion form of a system.

Figure 1 displays the response of Y_t to a shock S_t in panel (a) as well as the response of S_t to itself in panel (b). Panel (a) then breaks down the impulse response $\mathcal{R}_{s\to y}(h)$ into the response due to the internal propagation dynamics of Y_t, shown as a dashed red line; and the additional component coming from the persistence in S_t as subsequent changes in S_t propagate through the internal dynamics of Y_t. The sum of the two is the usual response, shown as a solid blue line. The response of S_t to its own shock is shown in panel (b).

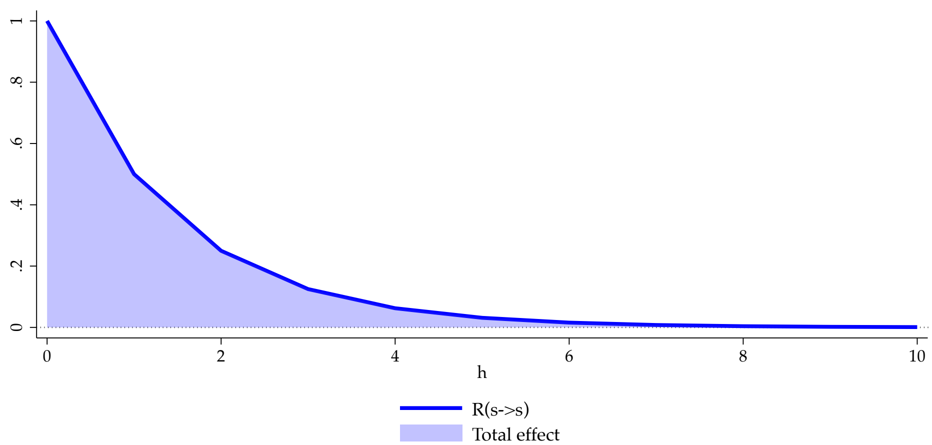
What is the researcher interested in? If the goal is to evaluate the effectiveness of modifications to the policy variable S_t, such as when one calculates a fiscal multiplier, it is clear that the usual representation of the impulse response presented in the solid blue line in panel (a) will overstate

Figure 1: *Decomposing an impulse response*

(a) *Response of Y_t to S_t*



(b) *Response of S_t to S_t*



Notes: The parameter choices of the model are: $\beta = 1$, $\rho = 0.5$, and $\psi = 0.75$. Panel (a), blue line is the usual impulse response $\mathcal{R}_{s\to y}(h)$ of Y_t in response to a shock in S_t. The dashed dotted line is the response that results from shutting down the subsequent response of S_t to its own shock. The shaded blue area represent the additional response generated by the persistence in S_t itself. Panel (b) is the response of S_t to itself, $\mathcal{R}_{s\to s}(h)$. See text.

the effect, an observation made, among others, by Mountford and Uhlig (2009) and Ramey (2011, 2016). Instead one can consider displaying *cumulative response ratios*, (Nath, Ramey, and Klenow, 2024), which I will denote $\mathcal{C}_h$.

Using Figure 1 as an illustration, the idea is to compare the area under the impulse response in panel (a) to the area under the impulse response in panel (b) to calculate the overall contribution of the shock to the response in Y_t relative to the overall change in S_t. That is:

$$\mathcal{C}_h = \frac{\frac{1}{h}\sum_{j=0}^{h} \mathcal{R}_{s\to y}(j)}{\frac{1}{h}\sum_{j=0}^{h} \mathcal{R}_{s\to s}(j)}. \tag{7}$$

Equation 7 can be seen as a ratio of two averages, the average effect of the shock on the outcome up to period h relative to the average effect of the shock to itself.

How can we estimate $\mathcal{C}_h$? Ramey (2016) and Jordà and Taylor (2025) suggest the following strategy. First, construct the following auxiliary variables:

$$\boldsymbol{y}_t(h) = \sum_{j=0}^{h} y_{t+j}; \qquad \boldsymbol{s}_t(h) = \sum_{j=0}^{h} s_{t+j}.$$

Next, let $\boldsymbol{z}_t$ denote a vector of instrumental variables for s_t which could simply be s_t if it is exogenous itself. Then note the moment condition:

$$E[\boldsymbol{z}_t(\boldsymbol{y}_t(h) - \mathcal{C}_h \boldsymbol{s}_t(h))] = 0,$$

which can be estimated by the generalized method of moments (GMM) or by simple ordinary least-squares (OLS) when s_t is exogenous and hence $\boldsymbol{z}_t = s_t$. The advantage of estimating $\mathcal{C}_h$ in this manner rather than estimating the responses first, cumulating them, and then taking the ratio is that it is much easier to obtain standard errors for $\mathcal{C}_h$. Finally note that if y_t and s_t are expressed in logs, then $\mathcal{C}_h$ can be interpreted similarly to an elasticity.

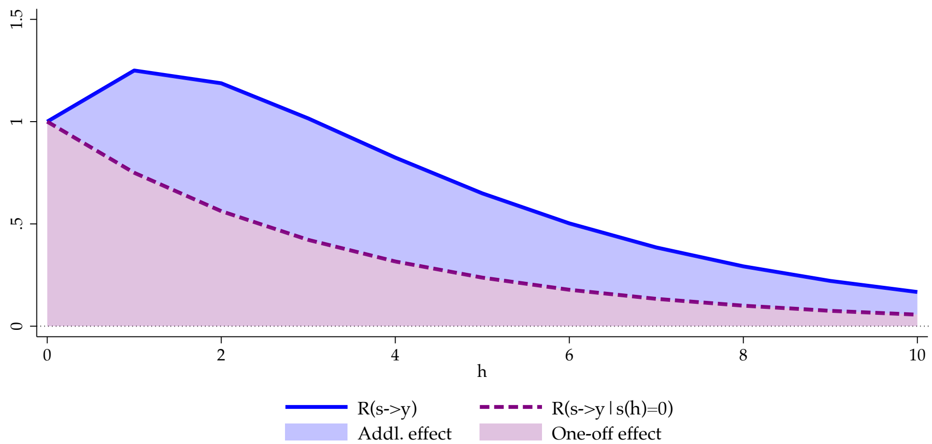
Like the previous section, it is important to focus on the question that the practitioner is interested in answering. Here, I show that in some instances, cumulative response ratios may be a more appropriate statistic than typical impulse responses. Again, local projections make this transition simpler than deriving the same ratios using a VAR. Thus, to a great extent, presenting impulse responses to characterize the dynamic properties of the data is heavily influenced by force of the VAR habit when perhaps other alternatives would, at times, be preferable.

3. DSGE models and lag length choices

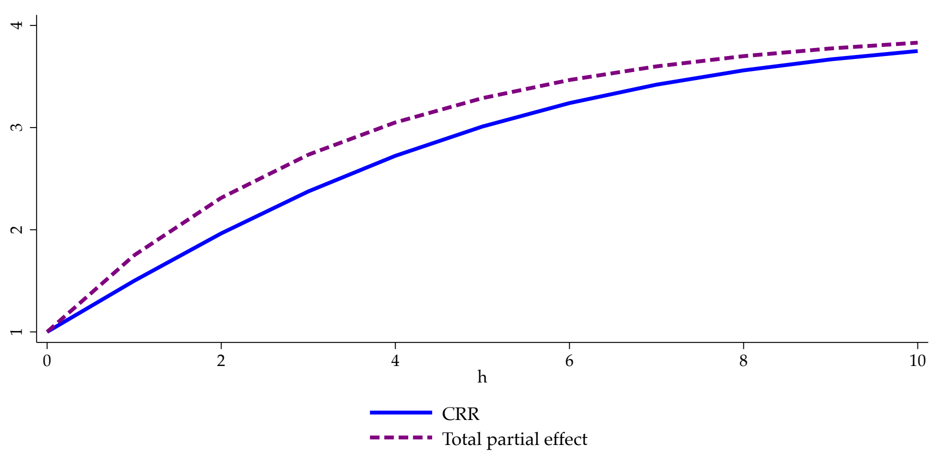
Let me return to the idea of comparing DSGE dynamics with those generated by a VAR. In this section I discuss the tension that exists when choosing the lag length. The discussion will help readers understand the prescriptions proposed by Montiel Olea et al. (2025) regarding this issue. I begin by noting that the log-linearized solution to many DSGE models can be cast as a state-space model (see, e.g. Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). In turn, the VAR representation from this state-space representation is, in most cases, a VAR(∞). For the purposes of the discussion, assume that this VAR(∞) is invertible so that one can derive the MA(∞) representation, itself the impulse response representation. Of course, invertibility is not always

Figure 2: *The cumulative response ratio*

(a) *Response of Y_t to S_t*



(b) *Comparing $\mathcal{C}_h$ with the true contribution of S_t*



Notes: Panel (a) replicates panel (a) in Figure 1. The red area under the dashed response is used in panel (b). Panel (b) plots the true cumulative contribution of S_t with a red dashed line and compares it to $\mathcal{C}_h$. See text.

guaranteed as Fernández-Villaverde et al. (2007) and others have argued, but this issue can be tabled for now.

In finite samples, one cannot estimate a VAR(∞), it has to be truncated at some lag p, which opens the question, to what extent is a VAR(p) a reasonable approximation to a VAR(∞)? Lewis and Reinsel (1985) provide a useful result in this respect. Write the VAR(∞) as $Y_t = A_1 Y_{t-1} + \ldots + \epsilon_t$ where I omit the constant for simplicity. Then, as long as $p \to \infty$ at a rate $p^3/T \to 0$ as the sample size, T, goes to infinity (and other more technical conditions), they show that:

$$\sqrt{T}[vec(\hat{A}_1'), \ldots, vec(\hat{A}_p') - vec(A_1'), \ldots, vec(A_p')] \to \mathcal{N}(0, \Sigma_a^*), \quad (8)$$

where Σ_a^* is the adjusted covariance matrix of the VAR coefficients that takes into account the truncation (the formula can be found, for example, in Lütkepohl, 2005).

Next, note that the impulse response (or MA) coefficients are a continuous transformation of the VAR coefficients and hence, by a version of Slutzky's theorem (and additional technical conditions), we have that:

$$\sqrt{T}[vec(\hat{B}_h') - vec(B_h)] \rightarrow \mathcal{N}(0, \Sigma_{b,h}) \quad \text{for} \quad h \leq p,$$

where again the formula for $\Sigma_{b,h}$ can be found in Lütkepohl (2005). The key takeaways here are that the truncation lag p has to be chosen at a rate $p^3/T \rightarrow 0$ and that consistency of the impulse response coefficients is only guaranteed for horizons smaller than p. This is the key tension. On one hand, asymptotic normality requires p not to grow too quickly, on the other hand, consistency of the impulse response coefficients is only guaranteed up to a horizon p.

A simple example clarifies these issues. In a sample of 100 observations and $p = 4$, the ratio p^3/T is 0.64, indicating that 4 lags may be too many as this ratio is close to 1 and not 0. On the other hand, restricting the lag length generates bias in the impulse response as the following example illustrates. Suppose the data are generated by a VAR(∞) but that we truncate at 1 lag (to make the calculations easier to grasp), then

$$\begin{array}{ll}
VAR(\infty) & VAR(1) \\
B_1 = A_1 & B_1^* = A_1 \\
B_2 = A_1^2 + A_2 & B_2^* = A_1^2 \\
B_3 = A_1^3 + 2A_1A_2 + A_3 & B_3^* = A_1^3 \\
B_4 = A_1^4 + \underbrace{3A_1^2A_2 + 2A_1A_3 + A_4}_{\substack{\text{potential bias from} \\ \text{omitted terms}}} & B_4^* = A_1^4 \\
\ldots & \ldots
\end{array} \tag{9}$$

Equation 9 reveals in stark contrast that truncating too soon can generate considerable bias in the estimation of impulse responses. In settings when more data is available, the previous discussion suggests that consistency of the impulse response will benefit from specifying longer rather than shorter lag lengths. Information criteria, a tool often used to determine the lag length in VARs and specially the Schwartz information criterion (SIC sometimes also referred as the Bayesian information criterion or BIC), put a strong premium on prediction accuracy, but this is not quite the right metric to reduce bias.

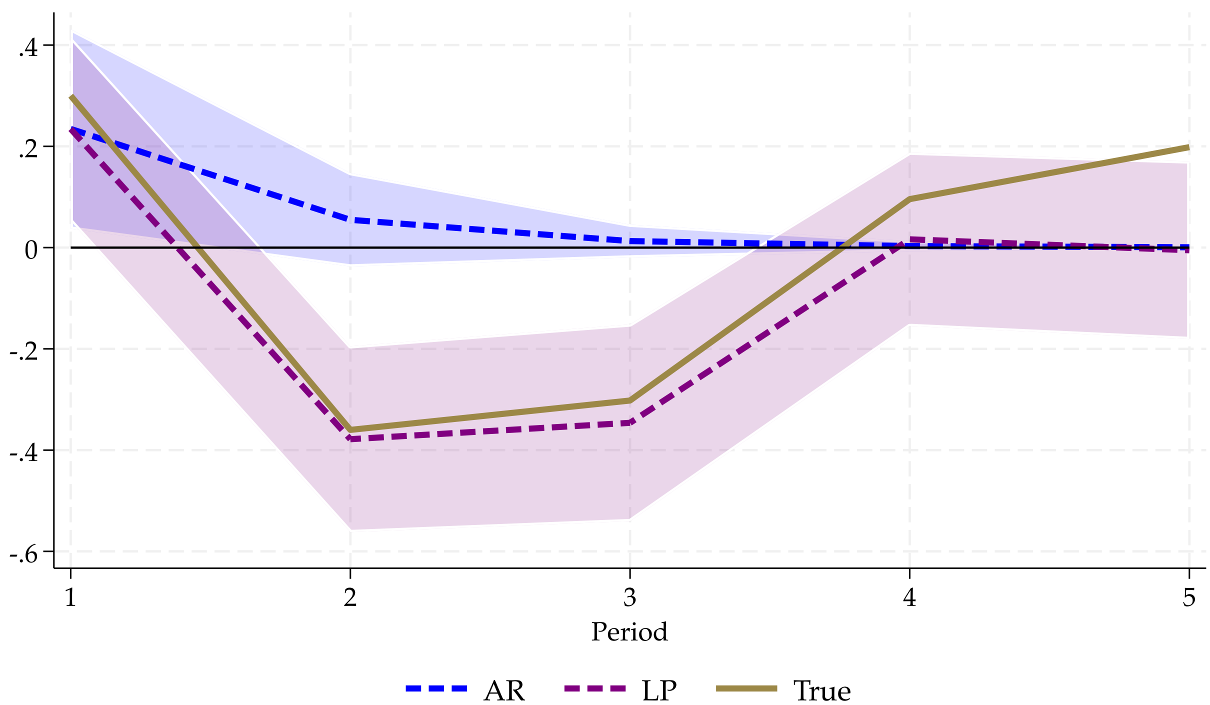
Thus, specifying VARs with longer lag lengths (if the sample size permits) will reduce bias. However, the efficiency gains from using a VAR relative to LPs are gradually eroded when specifying longer lags, as Montiel Olea et al. (2025) show, to the point that either method delivers the same response and error bands.

4. The bias-variance trade-off

When it comes to bias, one of the interesting features of LPs is their robustness to lag length misspecification. The intuition is easy to grasp. Each coefficient estimate is obtained from a different projection. The extent to which one specifies enough controls to make the residuals more or less white noise will ensure a low bias estimate. Of course, as Montiel Olea et al. (2025) discuss, one cannot be entirely careless as adding enough lags is still important.

Under similar conditions as discussed by Lewis and Reinsel (1985) for a VAR(∞), Jordà, Singh, and Taylor (2024) show that LPs provide consistent estimates of the impulse response at horizons beyond p, the lag truncation value, as long as h_{max} (the maximum horizon of the impulse response being considered) is not too large relative to the sample size.

Figure 3: *Example of bias-variance tradeoff between VARs and LPs*



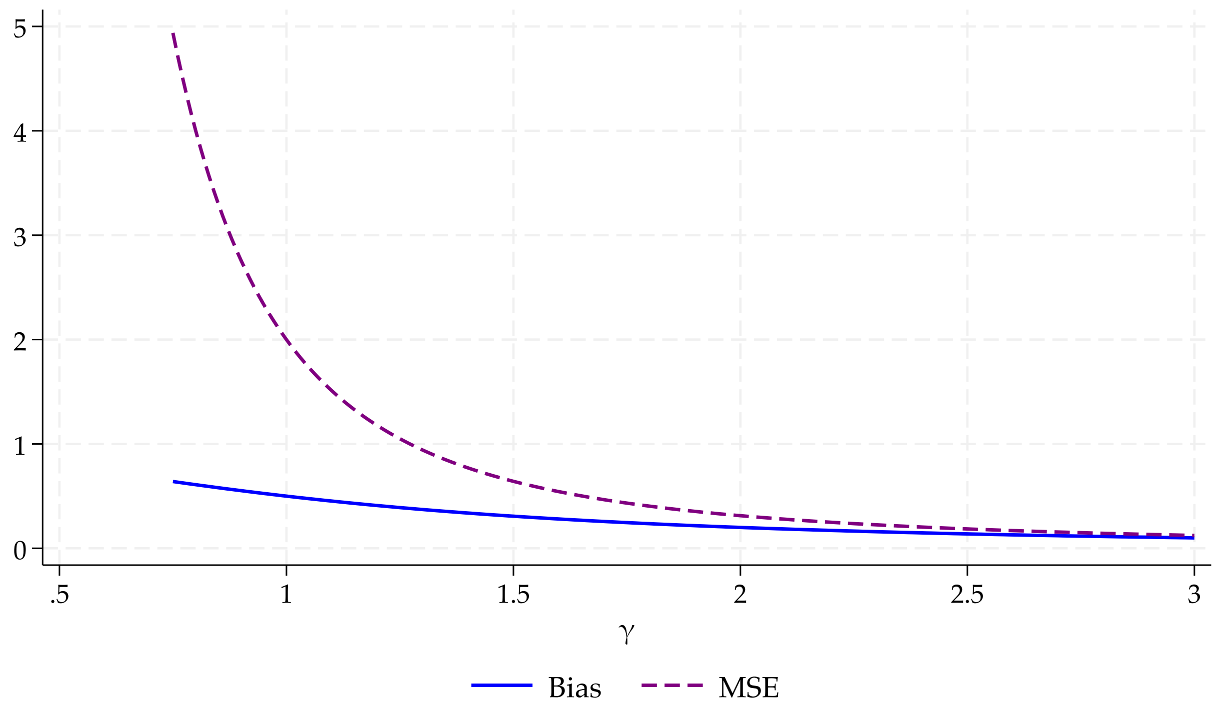
Notes: DGP is an AR(4) with lagged coefficients: $\rho_1 = 0.3$; $\rho_2 = -0.45$; $\rho_3 = -0.059$; $\rho_4 = 0.042$. Thus, the brown, solid line displays the true impulse response. The blue dashed line and the associated shaded area is the impulse response from an estimated AR(1) along with the 95% confidence region. The purple, dashed line and associated purple shaded region show the response calculated with LPs using one lag and their associated 95% confidence region. See text

Figure 3 provides a visual example of this dichotomy using a simple setup to more easily visualize the main issue. I simulated an AR(4) model with coefficients: $\rho_1 = 0.3$; $\rho_2 = -0.45$; $\rho_3 = -0.059$; $\rho_4 = 0.042$. using 1,000 burn-in observations and an estimation sample of 100 observations.

The figure shows the response estimates from an AR model with one lag versus LPs specified with one lag. The shaded areas are 95% confidence regions.

Figure 3 makes the point starkly. The AR(1) model is smooth, captures the first horizon almost perfectly, and has narrow error bands. The LPs capture the first 4 horizons quite precisely, albeit with wider error bands that get wider at longer horizons. I have generally seen this as a feature, not a bug. We should be more uncertain about the effect of an intervention in the more distant future, not less.

Figure 4: *The bias-variance tradeoff: OLS vs. IV*



Notes: Solid blue line shows the bias as a function of γ, the dashed red line shows the relative ratio of MSEs between IV and OLS. See text

This discussion presents practitioners with a choice not altogether unfamiliar. VARs will generally generate possibly biased responses but with a lower mean-squared error (MSE) than LPs, which tend to have lower bias, but are less efficient. And this issue is particularly problematic at horizons larger than the order of the VAR, p. VARs can badly miss features at medium and longer horizons if the lag length is short, as Figure 3 shows.

These trade-offs are similar to those facing researchers when estimating a regression by OLS versus IV. Generally, IV results are preferred because we value low bias over efficiency when we are after causal relations. I am thus left to wonder whether the same philosophy should be applied when considering whether to estimate impulse responses with a VAR or with LPs. Figure 3 shows that the response estimated with the AR(1) returns to zero after about two periods and the uncertainty vanishes at longer horizons. However, the true response is quite different, after the initial boost from the intervention, there is a negative response that last about 2-3 periods. The degree of certainty and the shape of the response of the AR(1) now appear to be quite deceiving.

A simple example illustrates the trade-offs in the OLS vs. IV case as a reminder of the general standards in applied practice. Suppose the DGP is $Y = \psi X + \epsilon$ with $X = \gamma Z + \epsilon$. Further assume that $V(Z) = V(\epsilon) = 1$. In this simple set-up, it is clear that as $\gamma \to 0$ we have both a weak instrument problem and higher bias when estimating by OLS. Going the other way, the larger γ is, the stronger the instrument and the lower the efficiency loss.

Figure 4 showcases these trade-offs by plotting the bias from OLS and the relative MSE of the IV estimator vs. the MSE of the OLS estimator. Take, for example, $\gamma = 1$. The bias is $1/2$ but the OLS estimator is twice as efficient as the IV estimator. For $\gamma = 0.75$, the bias is also 0.75 but now the IV MSE is about 5 times larger than the OLS MSE even though the signal to noise ratio has not deteriorated that much. In practice, however, this scenario would likely fall into the category of a ″weak″ instrument.

5. Conclusion

Montiel Olea et al. (2025) offer 12 lessons summarized into 8 recommendations. First, identification is separate from the choice of impulse response estimator. I would add, also subject to the object of interest as Sections 1 and 2 highlight. Not surprisingly, this continues to be an area of active research.

The next 5-6 recommendations share the same flavor, namely, they are about the tradeoffs between bias and variance of VARs and LPs. These are valuable and well sourced results that summarize findings from several well-known papers by the authors. In Sections 3 and 4 I tried to provide a different angle to put their recommendations in a different context for greater clarity, rather than to voice disagreement. Along the way, several important topics did not make the cut such as identification, applications to panel data, nonlinearities and other active areas of research. For this reason, I look forward to the second installment of the authors' review of this literature.

References

Angrist, Joshua D. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of business & economic statistics* 19(1): 2–28.

Bojinov, Iavor, and Neil Shephard. 2019. Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association* 114(528): 1665–1682.

Cloyne, James, Òscar Jordà, and Alan M. Taylor. 2023. State-Dependent Local Projections: Understanding Impulse Response Heterogeneity. NBER Working Paper 30971.

Cox, David R. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 20(2): 215–232.

Fernández-Villaverde, Jesús, Juan F Rubio-Ramírez, Thomas J Sargent, and Mark W Watson. 2007. ABCs (and Ds) of understanding VARs. *American economic review* 97(3): 1021–1026.

Frisch, Ragnar. 1933. Propagation Problems and Impulse Problems in Dynamic Economics 1. In *Business Cycle Theory, Part I Volume 4*, 305–336. Routledge.

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

Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor. 2024. The long-run effects of monetary policy. *Review of Economics and Statistics* forthcoming.

Jordà, Òscar, and Alan M. Taylor. 2025. Local Projections. *Journal of Economic LIterature* 63(1): 59–110.

Kolesár, Michal, and Mikkel Plagborg-Møller. 2025. Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly. *Journal of Business and Economic Statistics forthcoming*.

Kydland, Finn E, and Edward C Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society* 1345–1370.

Lewis, Richard, and Gregory C Reinsel. 1985. Prediction of multivariate time series by autoregressive model fitting. *Journal of multivariate analysis* 16(3): 393–411.

Long Jr, John B, and Charles I Plosser. 1983. Real business cycles. *Journal of political Economy* 91(1): 39–69.

Lucas, Robert E. 1976. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46.

Lütkepohl, Helmut. 2005. *New introduction to multiple time series analysis*. Berlin [u.a.]: Springer.

Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K Wolf. 2025. Local Projections or VARs? A Primer for Macroeconomists. In *NBER Macroeconomics Annual 2025: Volume 40*, volume 40. University of Chicago Press.

Mountford, Andrew, and Harald Uhlig. 2009. What are the effects of fiscal policy shocks? *Journal of Applied Econometrics* 24(6): 960–992.

Nath, Ishan B, Valerie A Ramey, and Peter J Klenow. 2024. How much will global warming cool global growth? Technical Report, National Bureau of Economic Research.

Rambachan, Ashesh, and Neil Shephard. 2019. Econometric analysis of potential outcomes time series: instruments, shocks, linearity and the causal response function. Unpublished. https://arxiv.org/abs/1903.01637.

Ramey, Valerie A. 2011. Identifying government spending shocks: It's all in the timing. *The quarterly journal of economics* 126(1): 1–50.

Ramey, Valerie A. 2016. Macroeconomic Shocks and Their Propagation. In *Handbook of Macroeconomics*, edited by Taylor, John B., and Harald Uhlig, volume 2, chapter 2, 71–162. Amsterdam: Elsevier.

Rubin, Donald B. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66(5): 688–701.

Rubin, Donald B. 1977. Assignment to treatment group on the basis of a covariate. *Journal of educational Statistics* 2(1): 1–26.

Rubin, Donald B. 1980. Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American statistical association* 75(371): 591–593.

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