ANNUAL REVIEWS

Annual Review of Economics

Local Projections for Applied Economics

Òscar Jordà[1,2,3]

[1]Federal Reserve Bank of San Francisco, San Francisco, California, USA

[2]Department of Economics, University of California, Davis, California, USA; email: ojorda@ucdavis.edu

[3]Centre for Economic Policy Research, London, United Kingdom

Annu. Rev. Econ. 2023. 15:607–31

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-082222-065846

JEL codes: C22, C23, C32, C33, E47

Keywords

local projections, vector autoregressions, panel data, potential outcomes

Abstract

The dynamic causal effect of an intervention on an outcome is of paramount interest to applied macro- and microeconomics research. However, this question has been generally approached differently by the two literatures. In making the transition from traditional time series methods to applied microeconometrics, local projections can serve as a natural bridge. Local projections can translate the familiar language of vector autoregressions and impulse responses into the language of potential outcomes and treatment effects. There are gains to be made by both literatures from greater integration of well-established methods in each. This review shows how to make these connections and points to potential areas of further research.

1. INTRODUCTION

Impulse response functions estimated with vector autoregressions (VARs) are a standard statistic used to investigate dynamic macroeconomic relationships. Though many associate impulse responses with the work of Sims (1980), there are references in economics already in the work of Frisch (1933). Of course, before their arrival to economics, impulse responses could trace their origin to the field of signal processing. A. W. Phillips (of the Phillips curve fame) built a hydromechanical analog computer in 1949 (know as the Monetary National Income Analogue Computer, or MONIAC) to illustrate the inner workings of Keynesian and Robertsonian economics. The effects of monetary policy were modulated by the flow of water through a system of pipes and valves representing different sectors of the economy, which activated a pen that drew an impulse response on a roll of graphing paper.

Traditionally, estimation of impulse responses has been viewed as a time series exercise that requires characterizing the entire dynamic system under consideration in order to study how policy interventions propagate over time, just as Phillips's MONIAC did. VARs were just a convenient and useful empirical approximation to such a dynamic system. Local projections (LPs) (Jordà 2005) shifted this system perspective to one where the impulse response could be directly estimated with univariate methods and without reference to other parts of the system.

LPs compare two conditional means of a future outcome given today's available information, one of which is subject to an intervention while the other is not. Immediately, one can think of this situation as comparing two forecasts under different circumstances or as comparing the conditional mean of treated and control subpopulations. Further, because forecasts and impulse responses are tightly linked, it quickly becomes apparent that traditional time series concepts and policy evaluation ideas stemming from the Rubin causal model (Rubin 1974) must be tightly connected as well. One area can benefit from the time series tradition of modeling dynamic relationships, as much as the other area can benefit from a rich tradition in the identification of causal effects. Though seemingly obvious, the connection took some time to sprout (see, e.g., Angrist & Kuersteiner 2011, Angrist et al. 2018), just as it took some time to make the connection between direct forecasts (see, e.g., Cox 1961, Klein 1968) and impulse responses (Jordà 2005).

The flexibility of LPs, which helps establish this macro-micro nexus, is at the same time a potential weakness. Because LPs are a univariate semiparametric approach, they cannot compete in mean-squared error terms with the specification of a traditional structural multivariate time series model (see, e.g., Plagborg-Møller & Wolf 2021, Li et al. 2022), even though in population they estimate the same response in many settings (again, see Plagborg-Møller & Wolf 2021). This should come as no surprise. The more restrictions one can place in describing the data, the more efficient the estimates, the smaller the mean-squared forecast errors, and the broader the scope to experiment with policy variations within the model. Moreover, since many models of the macroeconomy have solutions (or approximate solutions) that consist of a system of linear difference-differential equations, it is natural to impose the same structure on the data to extract estimates of the deep parameters of the model. LPs are not universally preferable, and one must recognize those situations where alternative methods have an edge.

However, by the same token, neither are traditional multivariate time series models universally preferable. For example, the consistency of an impulse response estimator depends on the truncation lag used to specify the infinite order approximation (see, e.g., Kuersteiner 2005, Jordà et al. 2020, Plagborg-Møller & Wolf 2021). This issue of potential misspecification is easily resolved using LPs (see, e.g., Jordà et al. 2020). Moreover, the natural efficiency losses of a less restrictive model, such as LPs, can often be significantly reduced, as several authors have shown (see, e.g., Barnichon & Brownlees 2019, Lusompa 2021, Montiel Olea & Plagborg-Møller 2021, Li et al.

2022). Furthermore, in infinite order settings, Xu (2023) shows that LPs are semiparametrically efficient if the order is allowed to grow with the sample. Importantly, just because a theoretical model of the economy is written in linear form, it does not mean that a structural linear multivariate model will describe the data correctly. More recently, the desire to stratify the responses according to some economic condition (see, e.g., Auerbach & Gorodnichenko 2012, Jordà & Taylor 2016, Tenreyro & Thwaites 2016, Ramey & Zubairy 2018) is trivially met using LPs, but it is much harder to meet using VARs. In general, nonlinearities can be investigated more easily in univariate rather than multivariate models.

The trade-off between VARs and LPs evokes that between ordinary least squares (OLS) and instrumental variables (IV) estimation. IV estimates are always less efficient (often times, wildly so), yet much of the profession prefers them to OLS estimates, almost regardless of the efficiency loss. The premium is on bias over efficiency, not on minimizing mean-squared error loss.[1] LPs by themselves do not resolve the issue of identification. However, researchers may prefer using LPs over VARs in settings where getting the dynamic response correctly is at a premium. More generally, efficiency losses in LPs can be greatly contained relative to the substantial bias improvements at medium to long horizons, especially with persistent data.

These issues become more pronounced as researchers tackle panel data and generally richer data sets. Moreover, the natural stratification resulting from the policy evaluation paradigm and the Kitagawa–Oaxaca–Blinder decomposition (Kitagawa 1955, Blinder 1973, Oaxaca 1973) does not fit traditional structural time series models well, whereas it is naturally accommodated using LPs (see Cloyne et al. 2023). Going the other way, policy evaluation of interventions that have effects over time, or interventions administered over time, possibly with different doses each time, could greatly benefit from the lessons learned over the past 40 years of applied macroeconomic research.

Extensions to panel data applications look like an especially fruitful area for LPs. In recent research, Dube et al. (2023) show that in difference-in-differences (DiD) settings with absorbing but heterogeneous treatments, LPs can greatly simplify the analysis and can even accommodate repeated treatments, thereby encompassing several of the methods recently proposed in the literature to tackle specific situations. Similar recent developments, such as regression discontinuity designs, probably deserve further exploration with LPs.

This review focuses on the applied macro-micro nexus through the method of LPs. The goal is not to provide an encyclopedic review of the LP literature but rather to highlight recent developments and avenues for research. The more points of commonality exist between these two venerable literatures, the more opportunities there are to advance each field through cross-pollination. The review therefore spends the first few sections going over basic estimation and inferential procedures for LPs and then dedicates the second half to showcasing LP applications that take advantage of widely used policy evaluation methods.

2. A BRIEF INTRODUCTION TO LOCAL PROJECTIONS

Let me begin by briefly discussing the intuition behind LPs with a simple example. Suppose $\boldsymbol{w}_t$ refers to a vector of stationary random variables observed over $t = 1, \ldots, T$ periods.[2] I assume stationarity for simplicity, although it is not necessary more generally.[3] Further assume that

[1] Though, admittedly, this is a trade-off worth revisiting.

[2] I use boldface to indicate vectors and capital letters for matrices.

[3] I am purposefully vague in the statement of many conditions to make the article more accessible. Formal statements can be found in the references provided.

$\boldsymbol{w}_t = (w_{1t}, \ldots, w_{jt}, \ldots, w_{kt})'$ for $j = 1, \ldots, k$ follows a simple VAR(1):

$$(\boldsymbol{w}_t - \boldsymbol{\mu}) = A(\boldsymbol{w}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t; \quad \boldsymbol{\epsilon}_t \sim D(\mathbf{0}, \Omega). \qquad 1.$$

It is well known that the response of w_{jt+h} due to a shock of size $\boldsymbol{\delta}_i$ in $\boldsymbol{\epsilon}_t$ is simply

$$\mathcal{R}_{ij}(h) = E[w_{jt+h}|\boldsymbol{\epsilon}_t = \boldsymbol{\delta}_i; \boldsymbol{w}_{t-1}] - E[w_{jt+h}|\boldsymbol{\epsilon}_t = \mathbf{0}; \boldsymbol{w}_{t-1}] = A^h_{[j,.]}\boldsymbol{\delta}_i,$$

for $h = 0, 1, \ldots, H$ and where $A^h_{[j,.]}$ denotes the j^{th} row of the matrix A raised to the h^{th} power. Here $\boldsymbol{\delta}_i$ refers to the size of the shock for each component of $\boldsymbol{\epsilon}_t$ chosen by the experimenter to reflect an identified experiment to the i^{th} variable in $\boldsymbol{w}_t$. Since the residuals $\boldsymbol{\epsilon}_t$ are usually correlated with one another, $\boldsymbol{\delta}_i$ can be seen as the linear combination that recovers the underlying structural residuals for the i^{th} variable. I set aside different ways to achieve identification (i.e., finding the right $\boldsymbol{\delta}_i$) to later sections. Finally, I use the notation $\mathcal{R}_{ij}(h)$ to denote the response from a shock in variable i to variable j that happens h periods after the initial intervention or shock.

Though this may seem like a restrictive example, note that the state-space representation of a VAR with p lags [a VAR(p)] and the approximate representation of other interesting stochastic processes have a VAR(1) representation. The more general case is derived by Jordà (2005). Further, for cointegrated systems readers may refer to Chong et al. (2012), who show how to decompose an impulse response in terms of the dynamics due to long-run equilibria and due to short-run dynamics separatedly. Jordà (2005) and Plagborg-Møller & Wolf (2021) formally establish the asymptotic equivalency of LPs and VARs under a variety of identification assumptions, and Stock & Watson (2018) and Plagborg-Møller & Wolf (2021) present the conditions under which IVs and LPs can be used to identify noninvertible[4] systems. Importantly, it is not necessary to assume that the data are generated by a VAR; it simply helps our reasoning.

As long as the model in Equation 1 accurately represents the data-generating process (DGP), a consistent estimate of the coefficient matrix A is all that is needed to calculate the impulse response at any horizon.

The LP approach instead uses recursive substitution,[5] yielding

$$(\boldsymbol{w}_{t+h} - \boldsymbol{\mu}) = A^{h+1}(\boldsymbol{w}_{t-1} - \boldsymbol{\mu}) + A^h\boldsymbol{\epsilon}_t + \cdots + A^0\boldsymbol{\epsilon}_{t+h},$$

with $A^0 = I$. The previous expression suggests that a regression of w_{jt+h} on $\boldsymbol{w}_{t-1}$ such as

$$w_{jt+h} = c_{jh} + \boldsymbol{\beta}_{jh+1}\boldsymbol{w}_{t-1} + v_{jt+h}; \quad \boldsymbol{v}_{t+h} = B_h\boldsymbol{\epsilon}_t + \cdots + B_0\boldsymbol{\epsilon}_{t+h} \qquad 2.$$

for $h = 0, 1, \ldots, H$ gives us an estimate of the impulse response, since we have

$$\mathcal{R}_{ij}(h) = E[w_{jt+h}|\boldsymbol{\epsilon}_t = \boldsymbol{\delta}_i; \boldsymbol{w}_{t-1}] - E[w_{jt+h}|\boldsymbol{\epsilon}_t = \mathbf{0}; \boldsymbol{w}_{t-1}] = \beta_{jh}\boldsymbol{\delta}_i,$$

which will be equal to $A^h_{[j,.]}\boldsymbol{\delta}_i$ as long as the DGP coincides with that in Equation 1. Note that $B_h = A^h$ in this simple example.

As discussed by Jordà (2005), estimates based on Equation 2 have several advantages, some of which are worth highlighting. First, Equation 2 can be estimated equation by equation, which makes estimation of nonlinearities and stratification simpler; examples of this are discussed in later sections. Second, Equation 2 is a direct estimate of the impulse response so that standard errors do not require the delta method or simulation-based methods (though they require adjusting for the serial correlation in the residuals or lag augmentation, as we shall see). Third, Equation 2 is less sensitive to misspecification since each impulse response coefficient is estimated using a different

[4]Loosely speaking, a noninvertible system is one in which the structural residuals cannot be recovered from the reduced-form residuals.
[5]Note that recursive substitution does not require stationarity.

regression. One can broadly think of LPs as a semiparametric approach to estimating impulse responses.

Moreover, when instruments are available, estimation of Equation 2 can be done with the method of IVs (see, e.g., Jordà & Taylor 2016, Ramey & Zubairy 2018). I postpone a more detailed discussion of IV estimation to Section 5, where I discuss how to estimate LPs generically using the generalized method of moments (GMM), and to Section 7, where I provide formal conditions for IV estimation. Here, though, there exist parallels with the literature on proxy VARs, where instruments are used to identify structural shocks from reduced-form shocks (see, e.g., Stock & Watson 2012, Mertens & Ravn 2013).

In summary, the main takeaways from this section are the following.

- Impulse responses can be equivalently estimated from a VAR or with LPs.
- However, because LPs can be estimated by univariate regression, they are more flexible, such as when generalizing to nonlinear settings and panel data.

3. TRANSFORMATIONS AND MULTIPLIERS

Macroeconomics data often exhibit trending behavior. Think of GDP or the price level over time, for example. Such trends can often be well described by a unit root—in time series parlance, they are I(1), or integrated of order one. If the data are log transformed, the first difference can be interpreted as the approximate percentage change in the variable (for example, the growth rate of GDP or the rate of inflation). To fix ideas, let y_t denote the log of an I(1) variable, let the first-difference be denoted as $\Delta y_t = y_t - y_{t-1}$, and let the long difference be denoted as $\Delta_h y_{t+h} = y_{t+h} - y_{t-1}$. The latter measures the approximate percentage change in the outcome from $t - 1$ to h periods in the future. In addition and for later use, let s_t denote a (randomly assigned) intervention of interest (to make things simple).

LPs can be estimated on the long differences ($\Delta_h y_{t+h}$) or the first differences (Δy_{t+h}) in response to an intervention s_t.[6] However, the interpretation of the impulse response is different in each case. LPs on $\Delta_h y_{t+h}$ measure the overall percentage change in the outcome since intervention. Notice that $\Delta_h y_{t+h} = y_{t+h} - y_{t+h-1} + y_{t+h-1} + \cdots - y_t + y_t - y_{t-1} = \Delta y_{t+h} + \cdots + \Delta y_t$. Adjusting the notation to indicate that s is the intervention that affects the outcome y however transformed, this means that the LP on the long difference measures the cumulative of the per-period percentage changes, that is, $\mathcal{R}_{sy}(h) = \mathcal{R}_{s\Delta_h y}(h) = \sum_{j=0}^{h} \mathcal{R}_{s\Delta y}(j)$.

A related statistic of interest is the multiplier. An early reference to the multiplier can be found in the work of Keynes (1936). The Keynesian (fiscal) multiplier compares two dynamic responses. The fiscal impetus in the first year a fiscal package is passed has effects on output that are felt over subsequent years. From this perspective, the multiplier might seem quite large. However, fiscal packages are usually implemented over several years, so that the overall effect of the fiscal package is best evaluated as the ratio of the overall gains in output relative to the overall fiscal expenditures over the duration of the package.

Therefore, the multiplier can be calculated as the sum of the cumulative changes in GDP due to the fiscal package over the cumulative sum of changes in the deficit due to the fiscal package. It is clear that the multiplier will be of interest in any setting in which an intervention is administered

[6]They can also be estimated on the levels, y_{t+h}. However, this is not generally recommended even though the response on the levels and the long differences coincide when y_{t-1} is included on the right-hand size. The reason is that when y_{t+h} is I(1), omitting y_{t-1} (as sometimes happens) can lead to invalid estimates and/or inference.

over several periods (we may call it a treatment plan) and one is interested in evaluating the overall effect of the treatment plan and not just the first intervention.

Consider a stripped-down model to fix ideas. Suppose that $y_t = \gamma s_t + u_t^y$ and $s_t = \rho s_{t-1} + u_t^s$, with $E(u_t^y, u_r^s) = 0$ for any value of r. In this simple model the treatment variable s_t is randomly assigned, though treatments are serially correlated. It is easy to see that $\mathcal{R}_{sy}(h) = \gamma\rho^h$ and $\mathcal{R}_{ss}(h) = \rho^h$. Define the multiplier as

$$m_h = \frac{\sum_{j=0}^{h} \mathcal{R}_{sy}(j)}{\sum_{j=0}^{h} \mathcal{R}_{ss}(j)} = \frac{\gamma \sum_{j=0}^{h} \rho^j}{\sum_{j=0}^{h} \rho^j} = \gamma. \qquad 3.$$

In economics terms, the overall effect of the treatment plan on the outcome happens to be the same as the effect on impact, though in more general settings this will not generally be the case. It is also useful to notice that since the effect on impact is γ and there are no internal propagation dynamics, the multiplier is simply the sum of the treatments over time, scaled by their per-period impact γ.

The LP estimator for $\mathcal{R}_{sy}(h)$ can be obtained from $y_{t+h} = \beta_h s_t + v_{t+h}^y$, and hence a direct estimate of $\sum_{j=0}^{h} \mathcal{R}_{sy}(j)$ can be obtained from the modified LP $Y_{t+h} = \theta_h^y s_t + v_{t+h}^y$, where $Y_{t+h} \equiv y_{t+h} + \cdots + y_t$. Clearly we have that $\theta_h^y = \beta_0^y + \cdots + \beta_h^y = \gamma(1 + \rho + \cdots + \rho^h)$. One can similarly construct $S_{t+h} = \theta_h^s s_t + v_{t+h}^s$ and therefore obtain $m_h = \theta_h^y / \theta_h^s$. However, one can go one step further by noting that if an instrument z_t is available such that $E(z_t, u_t^j) = 0$ for $j = y, s$, then we have

$$\begin{cases} \mathrm{cov}(Y_{t+h}, z_t) & = \theta_h^y \mathrm{cov}(s_t, z_t) \\ \mathrm{cov}(S_{t+h}, z_t) & = \theta_h^s \mathrm{cov}(s_t, z_t) \end{cases} \implies m_h = \frac{\theta_h^y}{\theta_h^s} = \frac{\mathrm{cov}(Y_{t+h}, z_t)}{\mathrm{cov}(S_{t+h}, z_t)},$$

which can be directly estimated from the auxiliary LPIV (LP estimated with IVs)

$$Y_{t+h} = m_h S_{t+h} + v_{t+h}, \qquad 4.$$

estimated using z_t as an IV. This is the approach proposed by Ramey (2016). The advantage of using this direct approach is that standard errors can be directly obtained from the regression output. Note that estimating this LP by OLS would not generate valid estimates even if S_{t+h} were completely assigned at random. In general, of course, we would include a vector of controls $\boldsymbol{x}_t$ in the previous expression and use more than one instrument if additional instruments are available. Ramey (2016) provides a more complete discussion.

In summary, this section has established the following.

- LPs estimated in levels or in long differences generate the same response (as long as enough lags of the outcome are included on the right-hand side).
- LPs can also be estimated in first differences, but the response then has a different economic interpretation.
- Multipliers evaluate the overall change in an outcome due to the overall treatment changes (treatment plan) over a given period of time.
- A direct estimate of such multipliers can be directly obtained with the auxiliary regression in Equation 4 estimated by IVs.

4. INFERENCE

As Equation 2 shows, the residuals of an LP generally have a moving average structure. Because they are dated t to $t + h$, they do not affect the consistency of the LP estimate, $\hat{\beta}_{jh}$. However, the residual serial correlation affects the construction of standard errors.

A semiparametric solution offered by Jordà (2005) was to use a Newey–West heteroscedasticity and autocorrelation consistent (HAC) estimator. Though this estimator is simple to use, several, more efficient alternatives have been proposed in the literature that are worth reviewing. Perhaps one of the more elegant solutions has been proposed by Montiel Olea & Plagborg-Møller (2021) and consists in adding an additional lag to the LP. Lag augmentation is known to improve inference in autoregressive models (see Toda & Yamamoto 1995, Dolado & Lütkepohl 1996, Inoue & Kilian 2020). A simple univariate example helps illustrate the main idea behind lag augmentation, though the method is shown to work with a generic VAR(p) DGP.

Thus, suppose the data are generated by a simple AR(1) model such as $w_t = aw_{t-1} + \epsilon_t$. For convenience, we may assume that w_t is strictly stationary with $|a| < 1$ and $\epsilon_t \sim D(0, \sigma^2)$. Consider estimating the LP $w_{t+h} = \beta_h w_t + v_{t+h}$. Plugging the AR(1) into the LP results in the expression $w_{t+h} = \beta_h \epsilon_t + \gamma_h w_{t-1} + v_{t+h}$.

Clearly, ϵ_t is not directly observable. However, we can use the Frisch–Waugh–Lovell logic to obtain β_h by regressing $(w_{t+h} - \gamma_h w_{t-1})$ on $(w_t - aw_{t-1})$. The estimator from this auxiliary two-step regression is such that

$$\hat{\beta}_h = \beta_h + \frac{\sum_{t=1}^{T-h} v_{t+h}\epsilon_t}{\sum_{t=1}^{T-h} \epsilon_t^2} \quad \rightarrow \quad \hat{\sigma}^2(\hat{\beta}_h) = \frac{\sum_{t=1}^{T-h} \hat{v}_{t+h}^2 \hat{\epsilon}_t^2}{(\sum_{t=1}^{T-h} \hat{\epsilon}_t^2)^2}.$$

The reason this approach works is that under the assumptions made on ϵ_t, the term $v_{t+h}\epsilon_t$ is serially uncorrelated even if v_{t+h} itself is serially correlated. This feature comes from the assumption that ϵ_t is strictly stationary with $E(\epsilon_t | \{\epsilon_s\}_{s \neq t})$.[7] As a result, a simple way to obtain correct inference for the LP is to add an additional lag as a regressor and then select a heteroscedasticty robust estimator to compute the standard errors—there is no longer a need to correct for serial correlation.[8] In the same paper, Montiel Olea & Plagborg-Møller (2021) propose a parametric wild bootstrap procedure where data are simulated from a VAR and then LPs are fitted to the simulated data to construct percentile t-confidence intervals.

A second option is to use a parametric specification of the residual covariance matrix. For example, Lusompa (2021) provides a simple feasible generalized least squares (FGLS) procedure that takes advantage of previous LP stages to correct the h^{th} stage residuals. Specifically, the idea is to estimate the first LP (i.e., for $h = 0$) as usual and collect for use in subsequent stages the residuals $\{\hat{\epsilon}_t\}$ and the estimate of the impulse response coefficient, say $\hat{\beta}_0$. For $h = 1$ construct the left-hand side variable $\tilde{y}_{t+1} = y_{t+1} - \hat{\beta}_0 \hat{\epsilon}_t$ and obtain $\hat{\beta}_1$ from the LP based on $\tilde{y}_{t+1}$. For h $= 2$, construct the left-hand side variable as $\tilde{y}_{t+2} = y_{t+2} - (\hat{\beta}_0 \hat{\epsilon}_{t+1} + \hat{\beta}_1 \hat{\epsilon}_t)$. Similar adjustments to the left-hand side variable are applied with subsequent horizons.

Lusompa (2021) shows that it is not necessary for the DGP to be a VAR for this procedure to correct for residual serial correlation (as long as the data are strictly stationary). Monte Carlo evidence in his paper shows that FGLS generates considerable gains in efficiency, especially when the data are highly persistent. Moreover, Lusompa (2021) also provides a bootstrap version using the score wild bootstrap (see Kline & Santos 2012) and a version for structural multi-step inference. Below I set up the GMM version of the LP estimator, which will make the underpinnings of this procedure perhaps easier to understand.

---

[7] In practical settings, it will be important to include enough lags to ensure that this condition is met.

[8] For example, in STATA, one would simply use the command `regress f‘h’.w w l(1/‘p’).w, vce(robust)`, where `‘h’` is a local that controls the impulse response horizon and `‘p’` is a local for the lag length. Other heteroscedasticity robust alternatives provided in STATA, such as `vce(hc2)` and `vce(hc3)`, may be preferred.

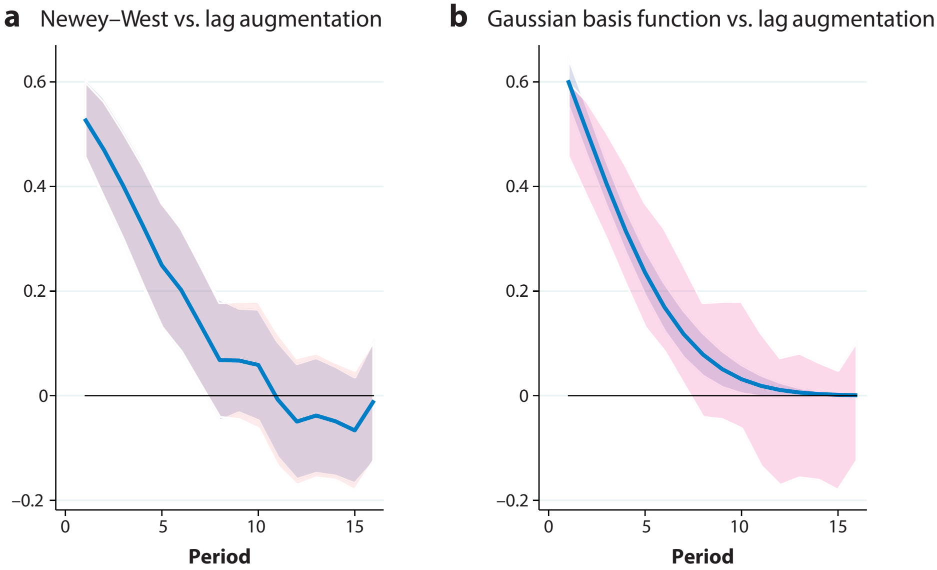

Figure 1

Comparing error bands. The figure shows the local projection of the response of x to y in the system

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix} + \begin{pmatrix} u_{yt} + 0.8u_{xt} \\ u_{xt} \end{pmatrix}; \quad u_{yt}, u_{xt} \sim N(0, 1).$$

Panel *a* compares error bands computed by Newey–West (6 lags) and lag augmentation. Panel *b* compares the error bands from panel *a* to a fitted Gaussian basis function estimated using the generalized method of moments with Newey–West robust standard errors.

Recent work by Xu (2023) shows that in settings where the true lag order is unknown and possibly infinite, LPs are semiparametrically efficient as long as controlled lags are allowed to grow with the sample size. This means that the efficiency loss relative to VARs diminishes the more lags one includes, and it effectively vanishes in the limit. The author then proposes two robust methods of inference.

Yet a fourth option consists of shrinking the variation of the LP coefficients. By adding some mild constraints, one can make considerable efficiency improvements yet retain much of the flexibility of LPs. A VAR does this automatically; with LPs this can be done in a variety of ways (see, e.g., Barnichon & Matthes 2018, Barnichon & Brownlees 2019, Miranda-Agrippino & Ricco 2021).

As an illustration, **Figure 1*a*** compares the error bands computed with Newey–West and lag augmentation using simulated data. **Figure 1*b*** fits a Gaussian basis function instead [such as the one proposed by Barnichon & Matthes (2018)] and shows error bands constructed using a direct GMM estimation with Newey–West robust standard errors. Panel *a* shows that, for this example, Newey–West and lag augmentation generate very similar (nearly indistinguishable) bands, as the theory predicts. Panel *b* shows that smoothing the LP responses can generate considerable reductions in uncertainty.

Several bootstrap methods have been proposed in the literature. The basic idea is as follows. First, estimate a VAR and generate bootstrap replicates of the data with it. Second, estimate LPs on these bootstrap replicates. The bootstrap sample of LPs can be used to construct inference. The procedure can be paired with a wild bootstrap to correct for potential heteroscedasticity. The reader should consult Montiel Olea & Plagborg-Møller (2019, 2021) for a detailed presentation of the procedures.

This section can be summarized as follows.

- Standard errors for LPs can be calculated using a HAC robust estimator.
- A simpler alternative is to use lag augmentation and a heteroscedasticity robust estimator.
- Smoothing the LP response can tighten inference considerably.
- Bootstrap procedures are available when simulation methods are preferred.

5. JOINT LOCAL PROJECTION ESTIMATION USING THE GENERALIZED METHOD OF MOMENTS

A useful way to think about estimation of LPs is by using the GMM, which naturally accommodates IV estimation. Let y_{t+h} be an outcome variable observed at time $t + h$, and let s_t be a treatment/intervention/policy variable whose effect on the outcome at some point in the future we are interested in characterizing. Let $\boldsymbol{x}_t$ refer to a $1 \times k$ vector of exogenous and predetermined variables that include lags of the outcome and the treatment variable. Let $\boldsymbol{z}_t$ denote a $1 \times l$ vector of instruments for s_t, which naturally include $\boldsymbol{x}_t$. When no instruments are available, then we have $\boldsymbol{z}_t = \boldsymbol{x}_t$, as one would have in a situation in which identification is achieved by conditioning on a rich set of right-hand-side variables via regression control.

Then, the h^{th} LP in a linear model satisfies the moment condition

$$E[g(\boldsymbol{w}_t; \beta_h)] = E[(y_{t+h} - \beta_h s_t)\boldsymbol{z}_t] = 0 \quad \text{for} \quad h = 0, 1, \ldots, H.$$

Stacking the left-hand-side variables into the $(H+1) \times 1$ vector $Y_t(H) = (y_t,\ y_{t+1},\ \ldots,\ y_{t+H})'$, the $(H+1) \times 1$ vector of impulse response coefficients $\boldsymbol{\beta} = (\beta_0,\ \beta_1,\ \ldots,\ \beta_H)'$ can be estimated as the solution to the GMM objective function

$$\max_{\boldsymbol{\beta}} \left(\frac{1}{T-H} \sum_{t=1}^{T-H} \boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\beta}) \right)' W_t \left(\frac{1}{T-H} \sum_{t=1}^{T-H} \boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\beta}) \right),$$

$$\max_{\boldsymbol{\beta}} \left(\frac{1}{T-H} \sum_{t=1}^{T-H} (Y_t(H) - \boldsymbol{\beta} s_t)\boldsymbol{z}_t \right)' W_t \left(\frac{1}{T-H} \sum_{t=1}^{T-H} (Y_t(H) - \boldsymbol{\beta} s_t)\boldsymbol{z}_t \right), \tag{5}$$

where, in the overidentified case, it is common to use the Newey–West version of the optimal weighting matrix W_t,

$$\hat{W}_t = \hat{\Gamma}_0 + \sum_{v=1}^{q} \left[1 - \frac{v}{q+1} \right] \left(\hat{\Gamma}_v + \hat{\Gamma}_v' \right),$$

$$\hat{\Gamma}_v = \frac{1}{T-H} \sum_{t=v+1}^{T-H} \boldsymbol{g}(\boldsymbol{w}_t; \hat{\boldsymbol{\beta}}) \boldsymbol{g}(\boldsymbol{w}_{t-v}; \hat{\boldsymbol{\beta}})'.$$

One could also take advantage of the known structure of the residual correlation in an LP using, for example, the continuously updated estimator of Hansen et al. (1996). Importantly, under standard regularity assumptions, estimation with Equation 5 delivers an estimate of the covariance matrix for $\boldsymbol{\beta}$, say $\Sigma_{\boldsymbol{\beta}}$, which will turn out to have important uses to conduct simultaneous inference, as I will show next. **Figure 1*b*** combines the GMM expressions just presented and assumes that the coefficients of the impulse response (the $\boldsymbol{\beta}$ in the previous expression) can be well approximated by a Gaussian basis function (which only depends on three parameters), as shown by Barnichon & Matthes (2018).

Of course, we do not need to be limited by linearity, and below I explore some natural nonlinear extensions, but then care must be observed in interpreting the LP. A simplified example illustrates how this should be done in a nonlinear setting:

$$y_{t+h} = \alpha_h + \beta_h s_t + \gamma_h s_t^2 + \epsilon_{t+h}.$$

Note then that the impulse response is no longer unique: It will depend on the values of the benchmark (say, $s_t = s_0$) and the treatment (say, $s_t = s_1 = s_0 + \delta$). That is, we have

$$\mathcal{R}(h)[s_0, \delta] = (\beta_h s_1 + \gamma_h s_1^2) - (\beta_h s_0 + \gamma_h s_0^2) = \beta_h \delta + \underbrace{\gamma_h \delta^2 + \gamma_h s_0 \delta}_{\text{nonlinear terms}}.$$

In this and in many alternative nonlinear specifications (some of which I discuss below), LPs remain linear in parameters so that the GMM objective function can be easily set up and estimated.

This section can be summarized as follows.

- The covariance matrix of the LP response coefficients can be obtained by jointly estimating the LPs with GMM.
- Nonlinearities are easily accommodated (and in many cases the LPs remain linear in parameters); however, one should be careful when interpreting the parameter estimates.

6. JOINT INFERENCE AND SIGNIFICANCE BANDS

Impulse response plots typically include error bands around the response estimates to provide a measure of estimation uncertainty. However, they are often misused to make inferential statements about the shape of the impulse response. This is problematic because impulse response coefficients are correlated with one another, as Jordà (2009) pointed out. The problem is analogous to using individual t-ratios instead of a χ^2- or F-test to do a joint hypothesis test in linear regression with correlated regressors. This section tackles these issues by building on the inferential methods presented above and building appropriate simultaneous inference bands. Ultimately, we want to display error bands that allow us to produce valid inferential statements under a variety of scenarios.

6.1. Simultaneous Inference

Asymptotically, and quite generally, we may assume that a vector of impulse responses estimates $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_H)'$ is such that $\hat{\boldsymbol{\beta}} \to \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta)$. The joint null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ could be tested with the traditional Wald statistic based on the Mahalonobis distance (Mahalonobis 1936), which turns out to be the sum of the square of the t-ratios when standardizing $\boldsymbol{\beta}$ by Σ_β. This Wald statistic will have an asymptotic χ^2 distribution with critical value $d(H, \alpha)$. Jordà (2009) then proposed constructing the individual critical values for the confidence interval of each β_h by using Scheffé's S-method (see Scheffé 1953), which in this example turns out to be $\sqrt{d(H, \alpha)/H}$. That is, Scheffé's S-method leads to more conservative error bands, which however have the correct coverage for any hypothesis test of the impulse response that can be expressed in linear form (such as joint significance).

Montiel Olea & Plagborg-Møller (2019) provide a more elegant solution. The idea is to provide bounds that can accommodate a variety of hypotheses of interest while providing the desired nominal coverage—say, with at least probability $1 - \alpha$. The idea is to construct an interval for each element in the response vector such that, in the worst case scenario, the null hypothesis of the element that is farthest from the estimate will still have the desired nominal coverage $1 - \alpha$. This is called the *sup-t* procedure and Montiel Olea & Plagborg-Møller (2019) show that it provides tighter bounds than the Scheffé S-method.

Here is how it works. Suppose as before that the estimates of an impulse response of interest are such that $\hat{\boldsymbol{\beta}} \to \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta)$. This asymptotic argument can be justified under a variety of rather general assumptions that apply to most situations observed in practice. Hence, define the auxiliary vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_H)' \sim \mathcal{N}(\mathbf{0}, \Sigma_\beta)$, with $\sigma_h = \Sigma_{[h,h]}$. The idea is to find the smallest critical value c such that for the collection of intervals around the response estimates we have that

$$P\left(\bigcap_{h=1}^{H}\left[\beta_h \in \hat{\beta}_h \pm c\,\hat{\sigma}_h\right]\right) \to P\left(\max_h |\sigma_h\,\eta_h| \leq c\right).$$

Alas, there are no tabulated values for the distribution of the maximum element of a normally distributed vector (the right-hand side of the previous expression), so that critical values have to be constructed via Monte Carlo simulation as

$$c = q_{1-\alpha}(\Sigma) \equiv q_{1-\alpha}\left(\max_h |\sigma_h^{-1}\,\eta_h|\right).$$

Based on this principle, Montiel Olea & Plagborg-Møller (2019) provide in addition bootstrap and Bayesian methods. The main advantage of constructing error bands in this manner is that inference on a subset of impulse response coefficients (e.g., are coefficients for horizons 3 to 6 different from zero?) will be correct. Of course, this comes at the cost of more conservative bands.

6.2. Significance

In many applications, it is common to see an impulse response with error bands that straddle the zero line. Many authors therefore conclude that the response is not significant, even though in many of these situations the response is uniformly positive (or negative). An example is provided in **Figure 2**. The figure shows the response of 100 times the log consumer price index (CPI) in the United States to a Romer & Romer (2004) shock over the sample 1969:Q1–2007:Q4. The

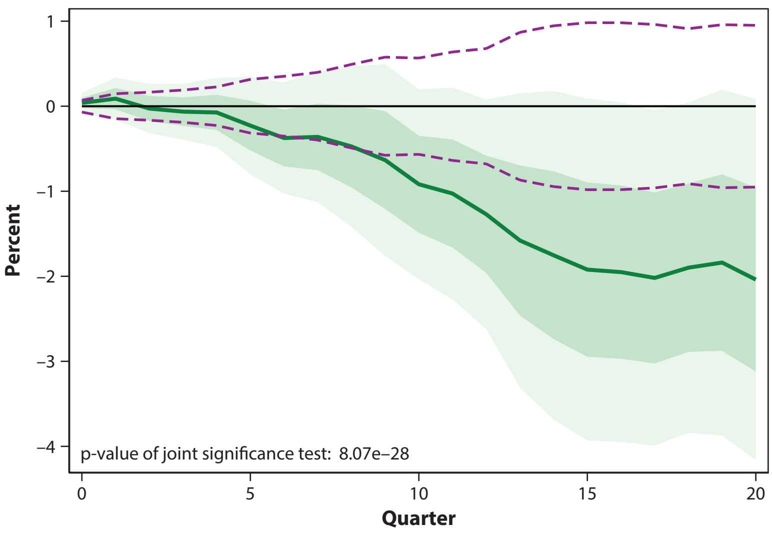

Figure 2

Response of inflation to a Romer & Romer (2004) monetary shock. The figure shows the local projection of the cumulative change of the consumer price index (CPI) on four lags of CPI inflation, four lags of real GDP growth, and four lags of the federal funds rate. The sample is the period 1969:Q1–2007:Q4. The shaded area shows two standard deviation pointwise confidence bands using heteroscedasticity robust standard errors. Dashed lines are computed by inverting the F-statistic around zero using Scheffé's method.

specification includes four lags of CPI inflation, real GDP growth, the federal funds rate, and the Romer & Romer shock itself.

The impulse response displayed in **Figure 2** is typical of many applications. It shows a time profile that is zero for about one year and is negative over the remaining four years. The point-wise error bands (shown at 95% confidence level) straddle the zero line, thus leading many researchers to conclude that the impulse response is not significant. However, as the figure shows, a joint test of the null hypothesis that all the response coefficients are zero can be easily rejected (with a p-value of 8.07e-28). To make the point clearer, **Figure 2** also displays two dashed lines. These are calculated by inverting the statistic of the null hypothesis that all impulse response coefficients are zero. That is, I display approximate 95% significance bands constructed as $\pm\hat{\sigma}_h\sqrt{d(H,\alpha)/H}$, since under the null hypothesis the coefficient estimates are approximately uncorrelated. Note that in the figure, for about the first two years, the impulse response is largely within these significance bands but clearly strays outside thereafter, thus confirming the result of the p-value (8.07e-28) reported for the joint test of significance.

Economically, the impulse response displayed shows that the CPI inflation is about 2 percentage points lower after five years, which is equivalent to a decline of CPI inflation of about 0.4% on average over the five years—a nonnegligible effect in economic terms even if individual response coefficients are imprecisely estimated.

What explains this disparity? As argued above, impulse response coefficients are highly correlated. Like regressions with near-collinearity, individual t-statistics are not significant, but a joint test of significance overwhelmingly rejects the null hypothesis. A good practice is therefore to report the joint test, which more closely corresponds to the scientific test of the hypothesis that the intervention has no effect on the outcome.

The arguments in this section can be summarized as follows.

- Error bands based on inverting t-ratios convey parameter estimation uncertainty but should not be used to conduct inference.
- The *sup-t* procedure provides correct coverage for multiple hypothesis testing but it is conservative.
- Assessing the overall significance of the response is best done with a joint test.

7. IDENTIFICATION

In a typical LP of the form

$$y_{t+h} = \alpha_h + \beta_h s_t + \boldsymbol{\gamma}_h \boldsymbol{x}_t + v_{t+h}; \quad h = 0, 1, \ldots, H,$$

with $\boldsymbol{x}_t$ containing exogenous and predetermined variables, OLS estimates will be consistent as long as variation in s_t is exogenous given $\boldsymbol{x}_t$. For example, the Cholesky identification assumption common in the VAR literature amounts to including the contemporaneous values of the system variables causally ordered first in $\boldsymbol{x}_t$. However, since the goal is to ensure that variation in s_t is as good as if it were exogenous, it seems that the safest route in general would be to include all available information to ensure orthogonality of s_t, regardless of the position of s_t in the causal order.

Similarly, identification with other methods common in the VAR literature (such as long-run identification restrictions or sign-based identification, for example) can be easily incorporated, as shown by Plagborg-Møller & Wolf (2021). I refer the reader to their paper for more details.

As previewed in the Introduction, one of the strong points of the policy evaluation literature is the emphasis on causation and hence on providing additional ways to approach identification. Expanding the idea behind regression control, I discuss in Section 8.4 inverse propensity score weighting, where control for $\boldsymbol{x}_t$ is allowed to be semiparametric (see, e.g., Hirano et al. 2003, Jordà & Taylor 2016, Angrist et al. 2018) based on ideas first discussed by Horvitz & Thompson (1952).

However, perhaps the more typical approach to identification in regression is the use of IVs, which can control for endogeneity. I have sprinkled references to identification with IVs at several points in the previous sections, in particular when discussing how to estimate LPs using GMM in Section 5. That said, it is useful to state formal conditions for when this approach is appropriate.

Specifically, denote the vector of instruments $\boldsymbol{z}_t$ for the intervention variable s_t. Further denote $\boldsymbol{z}_t^P = \boldsymbol{z}_t - \mathcal{P}(\boldsymbol{z}_t|\boldsymbol{x}_t)$ and, similarly, $s_t^P = s_t - \mathcal{P}(s_t|\boldsymbol{x}_t)$, where $\mathcal{P}(w|v)$ means the projection of w onto v. The first condition is that the instruments must be relevant:

- Relevance: $E(s_t^P, \boldsymbol{z}_t^{P'}) \neq \mathbf{0}$.

Next, we need the instruments to be exogenous. The exogeneity condition in an LPs setting is slightly different than usual due to the dynamic structure of the problem:

- Lead-lag exogeneity: $E(v_{t+h}, \boldsymbol{z}_t^{P'}) = 0$ for all h.

Stock & Watson (2018) and Plagborg-Møller & Wolf (2021) discuss these conditions in greater detail and provide more formal statements, though the main thrust of what is needed for IV estimation is summarized by the relevance and lead-lag exogenenity conditions just presented. Just like one can show the equivalence between VAR and LP responses, one can also show the equivalence of structural VAR (SVAR)-IV and LPIV (see, e.g., Mertens & Montiel-Olea 2018).

I conclude this section with a brief statement about an advantage of LPs over VARs highlighted by these authors. It consists in noting that, although invertibility is necessary for proxy-VARs (as VARs identified using IVs are typically referred to), this condition is not required for LPs. Invertibility essentially means that the structural residuals can be recovered from the reduced-form residuals. The condition usually fails when the span of the reduced-form residuals is smaller than the span of the structural shocks, as is common, for example, in models with news about future shocks.

The arguments in this section can be summarized as follows.

- Identification methods used for VARs can also be used for LPs.
- However, IV estimation is a more natural way to achieve identification.
- In addition to relevance, IVs need to meet a lead-lag exogeneity condition.
- An advantage of LPs over VARs is that invertibility is not required for identification.

8. THE DYNAMIC AND STATIC EFFECTS OF A POLICY INTERVENTION

In order to draw a closer link to the policy evaluation literature, I draw from variations of a simple model involving an outcome and a binary intervention or treatment. Suppose that y is an outcome variable of interest and s is a latent variable that determines policy/treatment/intervention according to $I(s_t) = I_t \in \{0, 1\}$. I leave the rule implicit in I_t undefined for the moment, though a simple example would be $I_t = I(s_t > c)$ for some c, as is common when estimating a logit or probit model. In some settings, I take s to be directly observable. In such cases, clearly, $s \neq 0$ can be directly interpreted as the dose given to a treated unit. Further, suppose that

$$\begin{cases} y_t = \beta I_t + \rho_{yy} y_{t-1} + u_t^y \\ s_t = \rho_{sy} y_{t-1} + \rho_{ss} s_{t-1} + u_t^s \end{cases}; \qquad \boldsymbol{u}_t \sim D\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_y & 0 \\ 0 & \sigma_s \end{pmatrix}\right). \tag{6}$$

This expression has several useful features. First, it is written in structural form. The residuals u_t^y and u_t^s are orthogonal to each other (explaining the switch from the ϵ_t to the u_t notation). Second, the outcome variable and the policy variable allow for possible internal propagation dynamics. Third, interventions can be thought of as randomly assigned when $\rho_{sy} = 0$. When $\rho_{sy} \neq 0$, interventions are endogenously determined by previous outcome values. Several interesting cases can be studied using this simple model.

8.1. No Serial Correlation

If $\rho_{ij} = 0$ for $i, j = s, y$, there are no internal propagation dynamics. A sample of $t = 1, \ldots, T$ observations behaves like a cross section. Hence, the effect of an intervention is β on impact and zero thereafter. This effect can be estimated as we would in a cross section, that is, by taking the following difference in sample means:

$$\mathcal{R}_{Iy}(0) = \frac{1}{T_1}\sum_{t=1}^{T} y_t I_t - \frac{1}{T_0}\sum_{t=1}^{T} y_t(1 - I_t); \quad T_1 = \sum_{t=1}^{T} I_t; \quad T = T_0 + T_1; \quad I_t \in \{0, 1\}. \qquad 7.$$

Of course, this could be simply estimated with the LP consisting of regressing y_t on I_t. The estimate of the constant term would be the mean for the untreated units, and the coefficient on I_t would be the effect of the intervention, β. It is easy to recognize from these two expressions the parallels with how one would estimate the treatment effect in a randomized controlled trial.

Using the potential outcomes notation, one would conjecture that the observed data come from a mixture distribution of two unobserved latent variables, $y_t(1)$ for observations in the treated subpopulation and $y_t(0)$ for the control subpopulation. Specifically, the observed data are $y_t = y_t(1)I_t + y_t(0)(1 - I_t)$. Since $y_t(j)$ for $j = 0, 1$ are not directly observable for each element of the sample, a quantity of interest is usually the average treatment effect, defined as $\tau(0) = E[y_t(1) - y_t(0)]$, which under random assignment can be directly estimated with Equation 7.

8.2. Serial Correlation in the Outcome

Next, suppose that $\rho_{yy} = \rho \neq 0$, but $\rho_{sy} = \rho_{ss} = 0$. In that case, we have $\mathcal{R}_{Iy} = (\beta,\ \beta\rho,\ \ldots, \beta\rho^h,\ \ldots)'$; that is, the intervention β is propagated by the internal dynamics of the outcome, but the assignment of the intervention I_t is still random. In principle, we can use the same difference in means as in the previous expression. However, to improve the efficiency of the estimator, we would want to take advantage of regressing the outcome on y_{t-1} first, since in general we have

$$\begin{aligned} y_{t+h} &= \rho^h \beta I_t + \rho^{h+1} y_{t-1} + v_{t+h}, \\ v_{t+h} &= u_{t+h}^y + \rho u_{t+h-1}^y + \cdots + \rho^h u_t^y + \\ &\quad \underbrace{\beta I_{t+h} + \rho\beta I_{t+h-1} + \cdots + \rho^{h-1}\beta I_{t+1}}_{\text{future interventions}}, \end{aligned}$$

which is just the LP of y_{t+h} on I_t and y_{t-1}. The residuals contain terms associated with future interventions or shocks. Under our assumptions, these are as if randomly assigned, so they do not cause an inconsistency with the LP estimate of $\rho^h\beta$. However, if the I_{t+h} are observable (and under the maintained assumptions for this example), nothing prevents one from including them as regressors (by construction they are uncorrelated with the u_{t+j}^y for any j), so that the LP that one would estimate becomes

$$y_{t+h} = a_0 I_{t+h} + a_1 I_{t+h-1} + \cdots + a_h I_t + c_h y_{t-1} + v_{t+h}; \quad v_{t+h} \sim MA(h),$$

where $a_j = \rho^j \beta$ for $j = 0, \dots, h$, $c_h = \rho^{h+1}$, and the short-hand notation $MA(h)$ indicates that the residuals have a moving-average structure of order h. In this case, we can therefore estimate the impulse response with a single LP set for the desired length, that is, $\hat{\mathcal{R}}_{Iy} = (\hat{a}_0, \dots, \hat{a}_h)'$.

How would one approach estimating the average treatment effect in similar fashion to Equation 7? Note that for this example, one would be interested in $\tau(h) = E[y_{t+h}(1) - y_{t+h}(0)|\Lambda_{t+h}]$, where $\Lambda_{t+h} = I_{t+h}, \dots, I_{t+1}; y_{t-1}$, and, based on our example, y_{t-1} is a summary statistic for the effects of previous interventions. Conditioning on future treatments isolates the effect of the current treatment. Thus, let $y_{t+h|\Lambda}$ denote the value of y_{t+h} conditional on Λ_{t+h} (say, from a regression of y_{t+h} on Λ_{t+h}). Then an alternative estimate of the impulse response is

$$\mathcal{R}_{Iy}(h) = \frac{1}{N_1} \sum_{t=1}^{N} y_{t+h|\Lambda} I_t - \frac{1}{N_0} \sum_{t=1}^{N} y_{t+h|\Lambda} (1 - I_t) = \beta \rho^h,$$

with $N = T - h$, $N_1 = \sum_{t=1}^{N} I_t$, and $N = N_0 + N_1$. In a moment, the usefulness of this derivation will become apparent.

8.3. Serially Correlated Interventions

Suppose that $\rho_{ss} = \rho \neq 0$ but $\rho_{yy} = \rho_{sy} = 0$. In this case, interventions are serially correlated but still randomly assigned. When a unit receives an intervention, it is likely that it will receive interventions in the next few periods, since we have that $I_{t+h} = I(\rho^h s_t + u^s_{t+h} + \rho u^s_{t+h-1} + \dots + \rho^{h-1} u^s_t)$. Interventions are still as if they were randomly assigned; however, the usual LP in this case would include past values of the intervention as a right-hand-side variable. That is, we have

$$y_{t+h} = \beta_h I_t + \rho_h I_{t-1} + v_{t+h}; \quad v_{t+h} \sim MA(h); \quad \hat{\mathcal{R}}_{Iy}(h) = \hat{\beta}_h.$$

In light of the previous example, it is natural to ask why would one not be also including future values of the intervention as regressors, as is done in the definition of Λ_{t+h}. Recent papers from the DiD literature argue that LPs estimate the wrong object for this reason (see, e.g., De Chaisemartin & D'Haultfoeuille 2022). However, this is just a confusion about the object of interest. In a typical impulse response, the effect of the intervention accommodates the possibility that future interventions will be subsequently administered with some probability, as is the case when s_t is serially correlated. That is, the usual macroeconomics response answers the question of identifying the likely effect on the outcome of an intervention at time t, recognizing that the intervention itself generates subsequent interventions. This is the effect we most likely see in the data. However, conditioning on future interventions is also valid but answers a different question, that of the effect of a one-off intervention.

De Chaisemartin & D'Haultfoeuille (2022) and others are interested in the effect of the intervention in isolation from any subsequent potential intervention. This is an equally legitimate question to ask. Here, once again, we can make a connection to a literature in applied macroeconomics that studies the fiscal multiplier (see, e.g., Mountford & Uhlig 2009, Ramey 2016, Ramey & Zubairy 2018), as I will show.

That said, a key observation is worth noting. In panel data settings where treatment effects may be heterogeneous across units, the difference between these two approaches matters. In a traditional time series setting, an implicit yet critical assumption is that the effect of subsequent treatments is homogeneous, that is, the specific time that treatment is administered does not alter the treatment effect, all else being equal. In the burgeoning literature on DiD estimation (see Roth et al. 2022 for an overview), it is becoming standard to assume that treatment is heterogeneous. I will return to this issue below.

By the same token, an issue often overlooked in the DiD literature is the role of expectations. That is, in a setting where agents expect interventions to follow after the initial (and possibly randomly assigned) intervention, their behavior will take into account such an eventuality. Thus, conditioning on past information and on future treatments will not completely account for the effect of expectations, except in situations where agents are completely backward looking, for example. It seems safer to adopt instead the standard macroeconomics practice of reporting the impulse response without removing the effect of future interventions and focus on measuring multiplier effects, as discussed above.

Summarizing, when interventions are serially correlated, an intervention today will likely be followed by subsequent interventions. The traditional impulse response measures the effect on the outcome of the entire treatment plan, that is, of the intervention implemented today and the set of subsequent interventions expected to follow due to serial correlation. This is the effect we are likely to see in the data. Thus a practitioner may well be interested in calculating a multiplier consisting of the sum total of the effects of the intervention plan on the outcome over some horizon, divided by the sum total of the interventions over that same horizon, as is done in the calculation of m_h presented in Section 3.

The policy evaluation literature tends to simply focus on the effect of the initial intervention by sterilizing the effect of subsequent interventions. In our example, this is equivalent to the multiplier (though in richer settings this will not be exactly the same). Note that the impulse response can also be thought of as the interaction of the one-time intervention effect scaled by the serial correlation pattern in the intervention plan. In practice, the role of expectations is often ignored, which can make sterilization based on future interventions insufficient and the estimates therefore invalid—it is a trivial violation of the no anticipation assumption common in DiD studies.

8.4. Endogenous Assignment

This final example will allow us to briefly discuss inverse propensity score estimators based on some conditional ignorability assumption (to be stated momentarily), often found in applied microeconomics research. Suppose that the coefficient on y_{t-1} in the equation for s_t is nonzero, that is, $\rho_{sy} = \rho \neq 0$. For simplicity, assume that $\rho_{yy} = \rho_{ss} = 0$.

In the general case in which the parameters of Equation 6 are unrestricted, the impulse response can be calculated in several ways. First, notice that if $\rho_{sy} \neq 0$ then assignment is no longer random; it is determined by past values of the outcome. However, owing to the structure of the problem, assignment is as if it were random if one linearly controls for y_{t-1}. Hence, the response $\mathcal{R}_{Iy}(h)$ can be calculated with a typical LP of y_{t+h} on I_t and y_{t-1}. Alternatively, one can first regress y_{t+h} on y_{t-1} and then use the residuals from this regression, which we can call $y_{t+h|t-1}$, to compute the difference in means. We obtain

$$\mathcal{R}(h) = \frac{1}{N_1}\sum_{t=1}^{N} y_{t+h|t-1} I_t + \frac{1}{N_0}\sum_{t=1}^{N} y_{t+h|t-1}(1 - I_t),$$

where N, N_0, and N_1 are as defined above.

Yet a third alternative is to use inverse propensity score weighting. In situations where the experimenters are willing to assume that, conditional on observables, assignment is as good as if it were random, they may not be willing to assume that the relationship is linear. What are the experimenters to do? Here I follow two recent macroeconomic applications by Jordà & Taylor (2016) and Angrist et al. (2018).

Let $\hat{p}_t = p(I_t = 1|\mathcal{I}_t)$ denote the propensity score, where $\mathcal{I}_t$ refers to information available up to t (or, in the example, simply y_{t-1}). Angrist et al. (2018) propose a conditional independence, or selection on observables assumption, so that $y_{t+h}(i) \perp I_t | y_{t-1}$ for $i = 0, 1$, where $y_{t+h}(i)$ denote potential outcomes. The assumption basically says that, conditional on y_{t-1}, assignment to treatment is not influenced by the potential outcomes one may experience. However, by properties of the propensity score, this assumption can be rewritten as $y_{t+h}(i) \perp I_t | \hat{p}_t$. Using the results of Hirano et al. (2003), Jordà & Taylor (2016) and Angrist et al. (2018) show that a doubly robust estimate of the impulse response is

$$\mathcal{R}_{Iy}(h) = \frac{1}{N_1}\sum_{t=1}^{N} y_{t+h|t-1}\frac{I_t}{\hat{p}_t} + \frac{1}{N_0}\sum_{t=1}^{N} y_{t+h|t-1}\frac{(1-I_t)}{1-\hat{p}_t},$$

where N, N_0, and N_1 are as defined before. The doubly robust feature is reflected on the fact that the notation $y_{t+h|t-1}$ indicates a regression of the outcome on past information. That is, one controls via regression and via the propensity score. In practice, there are more efficient ways of estimating the model using doubly robust, inverse propensity score weighting. Importantly, standard errors should be adjusted for the first-stage estimation uncertainty in $\hat{p}_t$. Of course, this could be done with, for example, a paired bootstrap.

We can summarize as follows.

- Feedback and internal propagation dynamics imply that interventions have effects over several periods, which in turn affect the likelihood of future interventions.
- When expectations are rational, one cannot simply control for future interventions to calculate the thought experiment of a one-off intervention.
- Instead, a better practice is to calculate multipliers.

9. THE KITAGAWA DECOMPOSITION

In the previous examples, treatment/intervention is conveniently assigned at random, a situation rarely encountered in practice with observational data. With random assignment, covariates provide tighter, more efficient estimates of the treatment effect, but otherwise, whether they are included or not has no effect on bias. However, this view assumes that the influence of the covariates on the outcome remains impervious to treatment. This is implicitly assumed in a VAR. What if this assumption is wrong? What if the manner in which a covariate interacts with the outcome depends on whether treatment is administered or not?

In applied microeconomics, one can account for how covariates and treatment interact using a decomposition first proposed by sociologist Evelyn Kitagawa (Kitagawa 1955) and introduced to economics by Oaxaca (1973) and Blinder (1973). An extensive review of this decomposition is provided by Fortin et al. (2011). I hence refer to this as the Kitagawa decomposition. It turns out that the Kitagawa decomposition provides a natural way for thinking about how to stratify LPs and even estimate time-varying LPs while still using simple regression analysis. The results that I present next are based on work by Cloyne et al. (2023).

Let me start with a simple cross-sectional setting first, with as stripped down a notation as possible. Without loss of generality, one can write $y(j) = \mu_j + v_j$ for $j = 0, 1$, the two potential outcomes (1 for treated, 0 for control), and where $E(v_j) = 0$. Covariates introduce heterogeneity. A simple way to model this heterogeneity is by assuming that $v_j = (\boldsymbol{x} - E(\boldsymbol{x}))\boldsymbol{\gamma}_j + \epsilon_j$, which ensures that $E(v_j) = 0$ by assuming that $E(\epsilon_j) = 0$. Here, $\boldsymbol{x}$ refers to a vector of exogenous or predetermined variables, which could include lags of the outcome and the treatment variables. Using the same notation as above, let $I(s_t) = I_t \in \{0, 1\}$ denote the treatment indicator, which I will denote simply as I when the subscript is redundant to understand the main ideas. Hence, the average treatment effect (under linearity) can be written as

$$\begin{aligned} E[y(1)|I=1] - E[y(0)|I=0] &= E[E[y(1)|\boldsymbol{x}, I=1] - E[y(0)|\boldsymbol{x}, I=0]] \\ &= E[\mu_1 + E[\boldsymbol{x} - E(\boldsymbol{x})|I=1]\boldsymbol{\gamma}_1 + E(\epsilon_1|I=1)] \\ &\quad - E[\mu_0 + E[\boldsymbol{x} - E(\boldsymbol{x})|I=0]\boldsymbol{\gamma}_0 + E(\epsilon_1|I=0)]. \end{aligned}$$

Note that by assumption $E(\epsilon_j|I = j) = 0$ for $j = 0, 1$. Further, by adding and subtracting $E[\boldsymbol{x} - E(\boldsymbol{x})|I = 1]\boldsymbol{\gamma}_0$, the previous expression can be rearranged into

$$\begin{aligned} E[y(1)|I = 1] - E[y(0)|I = 0] = \mu_1 - \mu_0 & \\ + E[\boldsymbol{x} - E(\boldsymbol{x})|I = 1](\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_0) & \\ + (E[\boldsymbol{x} - E(\boldsymbol{x})|I = 1] - E[\boldsymbol{x} - E(\boldsymbol{x})|I = 0])\boldsymbol{\gamma}_0. & \end{aligned} \quad 8.$$

Equation 8 hence decomposes the effect of treatment into three components: (*a*) a direct effect coming from the difference in unconditional means between treated and control subpopulations; (*b*) an indirect effect due to differences in the manner the covariates affect the outcome, which leads to the natural hypothesis $H_0 : \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0$; and (*c*) a composition effect due to the fact that in small samples, random assignment is imperfect. A test of the balance condition—if assignment is truly random, the means of the covariates should be the same in the treated and control subpopulations—is therefore a test of the null hypothesis $H_0 : \mu_{\boldsymbol{x}}^1 = \mu_{\boldsymbol{x}}^0$.

Based on these standard derivations, Cloyne et al. (2023) show that, under fairly general assumptions, these three effects can be obtained from the augmented LP

$$y_{t+h} = \underbrace{\mu_0^h + (\boldsymbol{x}_t - \overline{\boldsymbol{x}})\boldsymbol{\gamma}_0^h + I_t\beta^h}_{\text{usual LP}} + \underbrace{I_t(\boldsymbol{x}_t - \overline{\boldsymbol{x}})\boldsymbol{\theta}^h}_{\text{interaction}} + v_{t+h} \quad 9.$$

for $h = 0, 1, \ldots, H$ and $t = h, \ldots, T$, where v_{t+h} is a residual term. Based on this linear regression in the parameters, note that one can calculate the following three elements of the Kitagawa decomposition:

$$\begin{aligned} \text{Direct effect:} \quad & \hat{\mu}_1^h - \hat{\mu}_0^h = \hat{\beta}^h; \\ \text{Indirect effect:} \quad & (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}})(\hat{\boldsymbol{\gamma}}_1^h - \hat{\boldsymbol{\gamma}}_0^h) = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}})\hat{\boldsymbol{\theta}}^h; \\ \text{Composition effect:} \quad & (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0)\hat{\boldsymbol{\gamma}}_0^h. \end{aligned}$$

Cloyne et al. (2023) highlight several interesting features of this decomposition. First, the LP in Equation 9 is still linear in the parameters and, therefore, very easy to estimate. Second, as Fortin et al. (2011) highlight, the decomposition is noncausal, unless the $\boldsymbol{x}_t$ are identified (say, using an instrument). As long as interventions are as if randomly assigned with respect to the stratification, this should not pose a problem.[9] Third, if the sample is balanced, then the composition effect will be approximately zero, since we have $\overline{\boldsymbol{x}}_0 \approx \overline{\boldsymbol{x}}_1 \approx \overline{\boldsymbol{x}}$, which provides an easy way to check for failure of identification. Fourth, on average, the indirect effect will be close to zero in a balanced sample, since, as we have just seen, $\overline{\boldsymbol{x}}_1 \approx \overline{\boldsymbol{x}}$. However, the indirect effect can be quite large for individual values of $\boldsymbol{x}_t$, which can easily deviate from $\overline{\boldsymbol{x}}$. Fifth, note that the interaction of the treatment with $\boldsymbol{x}_t$ means that for each value of $\boldsymbol{x}_t$ we obtain a different impulse response. In other words, we just made the impulse response time varying as long as $\boldsymbol{\theta}^h \neq \mathbf{0}$.

As an example, consider a highly stylized economy. Suppose y_t refers to an economic activity outcome (e.g., output growth); let x_t refer to a monetary policy stance (say, the difference between the policy rate and the natural rate), which for simplicity I assume to have mean zero; and let $I_t \in \{0, 1\}$ if the government implements a fiscal consolidation ($I_t = 1$) of size s_t. These variables are assumed to be generated by the following potential outcomes model,

$$\begin{aligned} y_t(0) &= \mu_0 + \gamma_0 x_t + u_t^y \qquad \text{if} \quad I_t = 0, \\ y_t(1) &= \mu_1 s_t + \gamma_1 s_t x_t + u_t^y \qquad \text{if} \quad I_t = 1, \end{aligned}$$

[9] Gonçalves et al. (2022), following Cloyne et al. (2023), emphasize this point.

that is, if there is a fiscal consolidation, $I_t = 1$, the effect is scaled by the size of the consolidation, s_t. Further assume that

$$s_t = \rho_s s_{t-1} + u_t^s,$$
$$x_t = \rho_x x_{t-1} + u_t^x,$$

so that the observed data are generated by

$$y_t = \mu_0 + \gamma_0 x_t + I_t s_t(\beta + \theta x_t) + u_t^y.$$

Both policy variables, x_t, and I_t (and, indeed, s_t) are determined at random, which is not how we expect an economy to operate, of course. In this setting, suppose that we are interested in the impulse response to a fiscal consolidation of size $s_t = 1$, given that the monetary policy stance is x_t. We obtain

$$\mathcal{R}_{sy}(h) = \beta\rho_s^h + \theta\rho_x^h x_t.$$

If the monetary stance is neutral ($x_t = 0$), or if $\gamma_0 = \gamma_1$, then we have $\mathcal{R}_{sy}(h) = \beta\rho_s^h$, the usual impulse response in a linear model. However, if the stance is not neutral ($x_t \neq 0$) and $\gamma_0 \neq \gamma_1$, then the term $\theta\rho_x^h x_t$ will modulate the initial response to a fiscal consolidation. It is easy to see that the LP in Equation 9 would deliver direct estimates with which to construct $\mathcal{R}_{sy}(h)$ for any value of x_t.

Figure 3 provides an example of the type of analysis that the Kitagawa decomposition allows. It is based on a simulation of the simple example we just discussed. Panel *a* shows that, for each observation in the sample, x_t will attain a different value, which in turn will accentuate or attenuate the response. This is the most visible for the response to s_t on impact. To reinforce this point, panel *b* shows the response when $x_t = 0$ (*solid blue line*) with two standard error confidence bands. The response is purposely designed to be almost 0. However, note that depending on the value of the second treatment variable, x_t, the response can be greatly accentuated (nearly 2 on impact) or greatly attenuated (nearly -2 instead).

At this point, it is helpful to draw the connections to the multiplier calculation reported above. Let's focus on the fiscal experiment when the monetary stance is neutral, that is, $x_t = 0$. In that case, we have $\mathcal{R}_{sy}(h) = \beta\rho_s^h$, and the multiplier m_h is easily seen to be the same as that calculated in Section 3, that is, $m_h = \beta$. However, when $x_t \neq 0$, then we have $\mathcal{R}_{sy}(h) = \beta\rho_s^h + \theta\rho_x^h x_t$ and in this case the multiplier is

$$m_h(x_t) = \frac{\sum_{j=0}^h \mathcal{R}_{sy}(j)}{\sum_{j=0}^h \mathcal{R}_{ss}(j)} = \frac{\sum_{j=0}^h \beta\rho_s^j + \theta\rho_x^j x_t}{\sum_{j=0}^h \rho_s^j} = \beta + \underbrace{\theta x_t \frac{(1 + \rho_x + \rho_x^2 + \cdots + \rho_x^h)}{(1 + \rho_s + \rho_s^2 + \cdots + \rho_s^h)}}_{K_h(x_t)}.$$

In other words, the earlier equivalency between the multiplier and the average treatment effect breaks down since now we have $m_h(x_t) = \beta + K_h(x_t)$, which is a function of x_t.

In summary, this section has presented the following arguments.

- An intervention can modify the manner in which the covariates affect the outcome. This results in nonlinear effects that depend on the value of the controls at the time of the intervention.
- The Kitagawa decomposition allows for a natural stratification of the response.
- The extended LP is linear in parameters, yet it generates time-varying impulse responses.

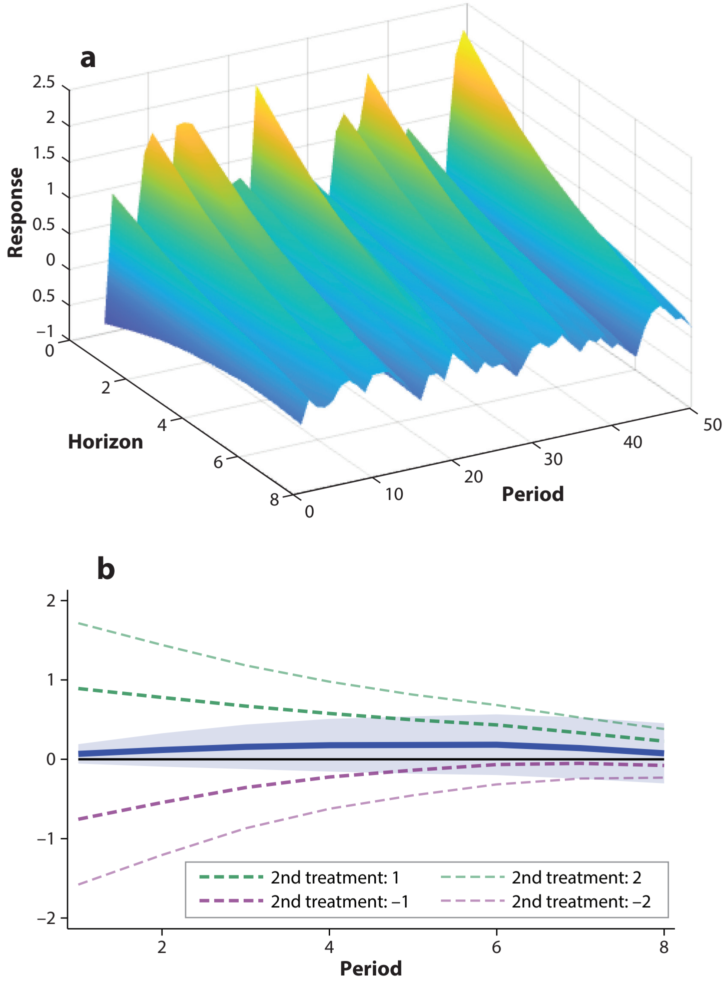

Figure 3

Kitagawa decomposition of the impulse response. Both panels in the figure are based on simulated data as in the example discussed in the text. Panel *a* displays the variation of selected impulse response coefficients over time due to variation in x_t. Panel *b* shows how the impulse response varies depending on different values of x_t.

10. PANEL DATA LOCAL PROJECTIONS

The ability to estimate impulse responses with univariate regression greatly facilitates their calculation in panel data settings. Given a sample of $i = 1, \ldots, n$ units observed over $t = 1, \ldots, T$ periods, the LP can be written as

$$y_{i,t+h} = \alpha_i + \delta_t + \beta_h s_{it} + \boldsymbol{\gamma}_h \boldsymbol{x}_{it} + v_{i,t+h},$$

where α_i are unit-fixed effects and δ_t are time-fixed effects. Because lags of the endogenous variable are often included in $\boldsymbol{x}_{it}$, potential incidental parameter biases could arise in short panels with a highly serially correlated endogenous variable (see, e.g., Álvarez & Arellano 2003). In such situations, an Arellano–Bond estimator or subsequent refinements are recommended (see, e.g., Arellano & Bond 1991, Arellano & Bover 1995, Blundell & Bond 1998).

Panels, in principle, offer opportunities to take advantage of the cross-sectional and time series dimensions to adjust standard errors for serial correlation and potential heteroscedasticity. Intuitively, clustering by unit/group uses the cross-sectional dimension to calculate autocovariances, thus adjusting for serial correlation nonparametrically and adjusting for clustering and heteroscedasticity. Clustering by time exploits the time dimension to construct residual variance estimates that vary by unit, thus correcting for heteroscedasticity nonparametrically.

However, the literature on clustered standard errors is rapidly evolving (see, e.g., Abadie et al. 2023). For example, Petersen (2009) emphasizes using clustering by unit rather than using Driscoll–Kraay standard errors (Driscoll & Kraay 1998)—that is, the panel version of a Newey–West standard error that emphasizes large T, small N asymptotics. Clustering by group relies on having a large number of groups so that the asymptotic approximation works in favor of clustering over Driscoll–Kraay. That said, Petersen (2009) finds the biases of Driscoll–Kraay to be relatively small in many situations. For short T panels, Monte Carlo evidence seems to indicate that it is sufficient to use time-fixed effects and one-way clustering.

Cameron et al. (2008, 2011) and Cameron & Miller (2015) emphasize that with small numbers of clusters, cluster-robust inference can be wildly incorrect (i.e., small N regardless of T asymptotics). In particular, simulation evidence by Cameron & Miller (2015) shows that there can be significant distortions, leading the authors to recommend bootstrap-based procedures (see also MacKinnon et al. 2022). Generally speaking, cluster-robust standard errors (and two-way clustering in particular) are highly sensitive to having a sufficient number of groups and time periods for the asymptotic theory to provide a good approximation.

To my knowledge, there is no theoretical result yet justifying lag augmentation procedures similar to those discussed above as a possible alternative/complement, though proving this result seems possible. As an example of an application of LPs in panels, I now discuss recent work by Dube et al. (2023).

It has been well documented (see, e.g., De Chaisemartin & D'Haultfoeuille 2020, Callaway & Sant'Anna 2021, Goodman-Bacon 2021, Sun & Abraham 2021) that in either static or distributed lag specifications in which there are multiple treated groups and treatment periods with heterogeneous treatment effects, the traditional two-way fixed effects (TWFE) estimator can be severely biased. This is true even when parallel trends hold with staggered treatment effects that are dynamic and possibly heterogeneous.

Previously treated units are invalid controls for currently treated units, which creates problems in distributed lag specifications. However, this is easily handled with LPs by using the clean control condition of Cengiz et al. (2019). In particular, let $P_t = 0$ for any period before intervention and 1 thereafter, and let $A_i = 0$ for an untreated unit, 1 for a treated unit. Hence, define $D_{it} = P_t \times A_i$. In a simple setting with no covariates, the DiD estimator of dynamic treatment effects can be estimated with

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h \Delta D_{it} + v_{i,t+h}; \quad h = 0, 1, \ldots, H$$

by restricting the sample to observations that are either

- treated: $\Delta D_{it} = 1$

or

- clean control: $\Delta D_{i,\,t+k} = 0$ for $k = -H, \ldots, h$.

The key advantage of LPs over distributed lag TWFE estimators is that differencing is in the outcomes, not in the treatments. Dube et al. (2023) show how the same estimator can be obtained by defining a dummy variable that is 1 for unclean controls by appropriately interacting this dummy variable with the regressors. I refer the reader to their paper for more details.

Simulation evidence shows that the LP approach is easier to implement, it is computationally much faster (which is important when using simulation-based inference, such as the bootstrap), and it provides consistent estimates of the treatment effects. Thus, this example shows that there are potentially many gains from incorporating LPs in other common situations in applied microeconomics in which treatment may have effects over more than one period. Examples may include regression discontinuity designs, synthetic control, and so on.

We can summarize as follows.

- Implementation of LPs on panel data is straightfoward.
- However, inference can be complicated by the dimensions of the panel. This is no different than in typical panels in microeconometrics research.
- LPs offer considerable advantages in the estimation of settings with dynamic, heterogeneous, staggered treatments.

11. CONCLUSION

In this review, I have focused on presenting the basic tools of LP estimation so as to establish the nexus between applications in applied macro- and applied microeconomics. That said, of necessity there is a great deal that fell on the editing floor. Examples include nonlinear applications of LPs, for example, to binary dependent data (see, e.g., Ferrari Minesso et al. 2023) and quantile regression (see, e.g., Jordà et al. 2022); Bayesian estimation of LPs (see, e.g., Tanaka 2020, Miranda-Agrippino & Ricco 2021); and smoothing and shrinkage methods (see, e.g., Barnichon & Matthes 2018, Barnichon & Brownlees 2019), to name a few.

More importantly, I have argued that impulse responses and dynamic treatment effects are close relatives to the point that LPs can offer a bridge between two literatures that up to this point appear to have developed quite separately from one another: time series analysis and methods in applied microeconomics research. The sections above showcase several examples where these literatures intersect. The hope is that this review will spur much more research at this intersection.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am grateful to Hélenè Rey for suggesting that I write this review and to my colleagues Colin Cameron, James Cloyne, Alan Taylor, and Takuya Ura for helpful suggestions.

LITERATURE CITED

Abadie A, Athey S, Imbens GW, Wooldridge J. 2023. When should you adjust standard errors for clustering? *Q. J. Econ.* 138(1):1–35

Álvarez J, Arellano M. 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71(4):1121–59

Angrist JD, Jordà Ò, Kuersteiner GM. 2018. Semiparametric estimates of monetary policy effects: string theory revisited. *J. Bus. Econ. Stat.* 36(3):371–87

Angrist JD, Kuersteiner GM. 2011. Causal effects of monetary shocks: semiparametric conditional independence tests with a multinomial propensity score. *Rev. Econ. Stat.* 93(3):725–47

Arellano M, Bond S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* 58(2):277–97

Arellano M, Bover O. 1995. Another look at the instrumental variable estimation of error-components models. *J. Econometr.* 68(1):29–51

Auerbach AJ, Gorodnichenko Y. 2012. Fiscal multipliers in recession and expansion. In *Fiscal Policy After the Financial Crisis*, ed. A Alesina, F Giavazzi, pp. 63–98. Chicago: Univ. Chicago Press

Barnichon R, Brownlees C. 2019. Impulse response estimation by smooth local projections. *Rev. Econ. Stat.* 101(3):522–30

Barnichon R, Matthes C. 2018. Functional approximations of impulse responses. *J. Monet. Econ.* 99:41–55

Blinder AS. 1973. Wage discrimination: reduced form and structural estimates. *J. Hum. Resourc.* 8(4):436–55

Blundell R, Bond S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *J. Econometr.* 87(1):115–43

Callaway B, Sant'Anna PH. 2021. Difference-in-differences with multiple time periods. *J. Econometr.* 225(2):200–30

Cameron AC, Gelbach JB, Miller DL. 2008. Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* 90(3):414–27

Cameron AC, Gelbach JB, Miller DL. 2011. Robust inference with multiway clustering. *J. Bus. Econ. Stat.* 29(2):238–49

Cameron AC, Miller DL. 2015. A practitioner's guide to cluster-robust inference. *J. Hum. Resourc.* 50(2):317–72

Cengiz D, Dube A, Lindner A, Zipperer B. 2019. The effect of minimum wages on low-wage jobs. *Q. J. Econ.* 134(3):1405–54

Chong Y, Jordà Ò, Taylor AM. 2012. The Harrod–Balassa–Samuelson hypothesis: real exchange rates and their long-run equilibrium. *Int. Econ. Rev.* 53(2):609–34

Cloyne JS, Jordà O, Taylor AM. 2023. *State-dependent local projections: understanding impulse response heterogeneity*. NBER Work. Pap. 30971

Cox DR. 1961. Prediction by exponentially weighted moving averages and related methods. *J. R. Stat. Soc. B* 23(2):414–22

De Chaisemartin C, D'Haultfoeuille X. 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *Am. Econ. Rev.* 110(9):2964–96

De Chaisemartin C, D'Haultfoeuille X. 2022. *Difference-in-differences estimators of intertemporal treatment effects*. NBER Work. Pap. 29873

Dolado JJ, Lütkepohl H. 1996. Making Wald tests work for cointegrated VAR systems. *Econometr. Rev.* 15(4):369–86

Driscoll JC, Kraay AC. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Rev. Econ. Stat.* 80(4):549–60

Dube A, Girardi D, Jordà O, Taylor AM. 2023. *A local projections approach to difference-in-differences event studies*. NBER Work. Pap. 31184

Ferrari Minesso M, Lebastard L, Le Mezo H. 2023. Text-based recession probabilities. *IMF Econ. Rev.* 71:415–38

Fortin N, Lemieux T, Firpo S. 2011. Decomposition methods in economics. In *Handbook of Labor Economics*, Vol. 4A, ed. O Ashenfelter, D Card, pp. 1–102. Amsterdam: Elsevier

Frisch R. 1933. Propagation problems and impulse problems in dynamic economics. In *Publikasjon*, Vol. 3, pp. 171–206. Oslo, Nor.: Univ. Økonomiske Inst.

Gonçalves S, Herrera AM, Kilian L, Pesavento E. 2022. *When do state-dependent local projections work?* Res. Dep. Work. Pap. 2205, Fed. Reserve Bank Dallas, Dallas, TX

Goodman-Bacon A. 2021. Difference-in-differences with variation in treatment timing. *J. Econometr.* 225(2):254–77

Hansen LP, Heaton J, Yaron A. 1996. Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Stat.* 14(3):262–80

Hirano K, Imbens GW, Ridder G. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71(4):1161–89

Horvitz DG, Thompson DJ. 1952. A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.* 47(260):663–85
Inoue A, Kilian L. 2020. The uniform validity of impulse response inference in autoregressions. *J. Econometr.* 215(2):450–72
Jordà Ò. 2005. Estimation and inference of impulse responses by local projections. *Am. Econ. Rev.* 95(1):161–82
Jordà Ò. 2009. Simultaneous confidence regions for impulse responses. *Rev. Econ. Stat.* 91(3):629–47
Jordà Ò, Kornejew M, Schularick M, Taylor AM. 2022. Zombies at large? Corporate debt overhang and the macroeconomy. *Rev. Financ. Stud.* 35(10):4561–86
Jordà Ò, Singh SR, Taylor AM. 2020. *The long-run effects of monetary policy*. NBER Work. Pap. 26666
Jordà Ò, Taylor AM. 2016. The time for austerity: estimating the average treatment effect of fiscal policy. *Econ. J.* 126(590):219–55
Keynes JM. 1936. The marginal propensity to consume and the multiplier. In *The General Theory of Employment, Interest, and Money*, pp. 101–16. London: Macmillan
Kitagawa EM. 1955. Components of a difference between two rates. *J. Am. Stat. Assoc.* 50(272):1168–94
Klein LR. 1968. *Essay on the Theory of Economic Prediction*. Helsinki, Finl.: Yrjo Jahnsson Lect.
Kline P, Santos A. 2012. A score based approach to wild bootstrap inference. *J. Econometr. Methods* 1(1):23–41
Kuersteiner GM. 2005. Automatic inference for infinite order vector autoregressions. *Econometr. Theory* 21(1):85–115
Li D, Plagborg-Møller M, Wolf CK. 2022. *Local projections vs. VARs: lessons from thousands of DGPs*. NBER Work. Pap. 30207
Lusompa AB. 2021. *Local projections, autocorrelation, and efficiency*. Unpublished manuscript, Fed. Reserve Bank Kansas City, Kansas City, MO
MacKinnon JG, Nielsen MØ, Webb MD. 2022. Cluster-robust inference: a guide to empirical practice. *J. Econometr.* 232(2):272–99
Mahalonobis P. 1936. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* 2:49–55
Mertens K, Montiel-Olea JL. 2018. Marginal tax rates and income: new time series evidence. *Q. J. Econ.* 133(4):1803–84
Mertens K, Ravn MO. 2013. The dynamic effects of personal and corporate income tax changes in the United States. *Am. Econ. Rev.* 103(4):1212–47
Miranda-Agrippino S, Ricco G. 2021. *Bayesian local projections*. Work. Pap., Univ. Warwick, Coventry, UK
Montiel Olea JL, Plagborg-Møller M. 2019. Simultaneous confidence bands: theory, implementation, and an application to SVARs. *J. Appl. Econometr.* 34(1):1–17
Montiel Olea JL, Plagborg-Møller M. 2021. Local projection inference is simpler and more robust than you think. *Econometrica* 89(4):1789–823
Mountford A, Uhlig H. 2009. What are the effects of fiscal policy shocks? *J. Appl. Econometr.* 24(6):960–92
Oaxaca R. 1973. Male-female wage differentials in urban labor markets. *Int. Econ. Rev.* 14(3):693–709
Petersen MA. 2009. Estimating standard errors in finance panel data sets: comparing approaches. *Rev. Financ. Stud.* 22(1):435–80
Plagborg-Møller M, Wolf CK. 2021. Local projections and VARs estimate the same impulse responses. *Econometrica* 89(2):955–80
Ramey V. 2016. Chapter 2 - Macroeconomic shocks and their propagation. In *Handbook of Macroeconomics*, ed. JB Taylor, H Uhlig, pp. 71–162. Amsterdam: Elsevier
Ramey VA, Zubairy S. 2018. Government spending multipliers in good times and in bad: evidence from US historical data. *J. Political Econ.* 126(2):850–901
Romer CD, Romer DH. 2004. A new measure of monetary shocks: derivation and implications. *Am. Econ. Rev.* 94(4):1055–84
Roth J, Sant'Anna PH, Bilinski A, Poe J. 2022. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. arXiv:2201.01194[econ.EM]
Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66(5):688–701
Scheffé H. 1953. A method for judging all contrasts in the analysis of variance. *Biometrika* 40(1–2):87–110
Sims CA. 1980. Macroeconomics and reality. *Econometrica* 48(1):1–48

Stock JH, Watson MW. 2012. Disentangling the channels of the 2007–09 recession. *Brook. Pap. Econ. Act.* 43(1):81–156

Stock JH, Watson MW. 2018. Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *Econ. J.* 128(610):917–48

Sun L, Abraham S. 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J. Econometr.* 225(2):175–99

Tanaka M. 2020. Bayesian inference of local projections with roughness penalty priors. *Comput. Econ.* 55(2):629–51

Tenreyro S, Thwaites G. 2016. Pushing on a string: US monetary policy is less powerful in recessions. *Am. Econ. J. Macroecon.* 8(4):43–74

Toda HY, Yamamoto T. 1995. Statistical inference in vector autoregressions with possibly integrated processes. *J. Econometr.* 66(1–2):225–50

Xu KL. 2023. *Local projection based inference under general conditions*. Tech. Rep., Indiana Univ., Bloomington, IN