

IMPULSE RESPONSE ESTIMATION BY SMOOTH LOCAL PROJECTIONS

Regis Barnichon and Christian Brownlees*

Abstract—Local projections (LP) is a popular methodology for the estimation of impulse responses (IR). Compared to the traditional VAR approach, LP allow for more flexible IR estimation by imposing weaker assumptions on the dynamics of the data. The nonparametric nature of LP comes at an efficiency cost, and in practice, the LP estimator may suffer from excessive variability. In this work, we propose an IR estimation methodology based on B-spline smoothing called smooth local projections (SLP). The SLP approach preserves the flexibility of standard LP, can substantially increase precision, and is straightforward to implement. A simulation study shows that SLP can deliver substantial gains in IR estimation over LP. We illustrate our technique by studying the effects of monetary shocks where we highlight how SLP can easily incorporate commonly employed structural identification strategies.

I. Introduction

IMPULSE response (IR) functions are a key tool for summarizing the dynamic effects of structural shocks on economic time series. While vector autoregressions (VAR) have been traditionally used to identify structural shocks and simultaneously recover the corresponding IRs, the rising popularity of the narrative identification approach popularized an alternative IR estimation approach: the local projections (LP) of Jordá (2005).¹

In its basic formulation, the LP approach consists of running a sequence of predictive regressions of a variable of interest on a structural shock for different prediction horizons. The IR is then obtained from the sequence of regression coefficients of the structural shock. This approach has a number of advantages in comparison to VARs: LP does not impose specific dynamics on the variables in the system, does not suffer from the curse of dimensionality inherent to VARs, and can more easily accommodate nonlinearities (Auerbach & Gorodnichenko, 2012). However, in the LP framework, the IR is expensively parameterized and the IR estimator can have a large variability (Ramey, 2012, 2016).

Received for publication January 3, 2017. Revision accepted for publication June 7, 2018. Editor: Yuriy Gorodnichenko.

*Barnichon: San Francisco Fed and CEPR; Brownlees: Universitat Pompeu Fabra and Barcelona GSE.

We thank Majid Al Sadoon, Mila Cheng, Jordi Galí, Felix Geiger, Oscar Jordà, Dennis Kristensen, Jaime Martínez-Martín, Barbara Rossi, Arthur Taburet, Andrea Tamoni, and seminar participants for helpful comments. C.B. acknowledges financial support from the Spanish Ministry of Science and Technology (grant MTM2015-67304-P), the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), and Fundación BBVA scientific research grant (PR16_DAT_0043) on analysis of big data in economics and financial applications. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of San Francisco or the Federal Reserve System. Any errors are our own. Matlab and R implementations of the procedures presented in this paper are available from the authors upon request to the authors.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00778.

¹For recent uses of the narrative identification approach, see Romer and Romer (2004, 2010) and Ramey and Zubairy (2018).

In this work, we introduce an IR estimation methodology, smooth local projections (SLP), that builds on penalized B-splines (Eilers & Marx, 1996). We model the sequence of IR coefficients as a linear combination of B-splines basis functions, and we estimate the coefficients of this linear combination using a shrinkage estimator that shrinks the IR toward a polynomial. SLP nest two important IR estimators. SLP coincide with LP when the degree of shrinkage is low and with an Almon (1965) polynomial distributed lag model when the degree of shrinkage is high. A cross-validation criterion is suggested to choose the degree of shrinkage between these two extremes.

SLP have a number of highlights. First, the methodology can substantially increase the estimation accuracy of LP while preserving flexibility. Second, SLP estimation boils down to standard ridge regression, which is straightforward to implement. Third, SLP, like standard LP, can be used to recover structural IRs in conjunction with a number of identification schemes (e.g., timing restrictions, instrumental variables).

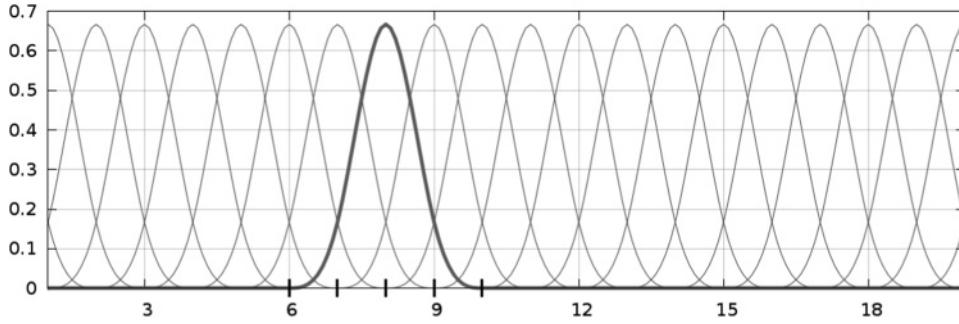
A simulation study is used to assess the finite sample performance of our proposed methodology. Results show that SLP delivers substantial improvements over LP or VARs for a range of DGPs calibrated to real data.

In this work, we focus on using SLP for IR point estimation. In empirical applications, IR confidence intervals are also a natural object of interest, and we propose a procedure for constructing IR confidence intervals using the SLP estimator. We do not study the theoretical properties of such a procedure, but the simulation study shows that our SLP confidence intervals perform similar to LP confidence intervals.

Finally, we illustrate our methodology by studying the effects of monetary shocks on GDP growth and inflation. For identification, we use both timing restrictions and an instrumental variable approach using the Romer and Romer (2004) narrative shocks as instrument. While the LP-based IRs can be erratic, the SLP-based IRs are more regular and easier to interpret.

Our paper contributes to a rapidly growing macroeconomic literature that relies on LP to estimate structural impulse responses (Ramey, 2016). LP can be seen as a modern offshoot of the distributed lag (DL) literature (Sims, 1974), and SLP can be seen as a modern version of Shiller's (1973) smoothness priors for DL models. More recently, a number of working papers have proposed related strategies to obtain smoother or regularized estimates of the IR (among others, Barnichon & Matthes, 2018, and Miranda-Agrippino & Ricco, 2017), but our approach has the advantage of being as straightforward to implement as LP. Although not based on B-splines smoothing, a complementary paper is by Plagborg-Møller (2016), who provides methods for optimally selecting the degree of smoothing and constructing confidence bands. Finally, our paper can be cast into the broader context

FIGURE 1.—B-SPLINE BASIS



The figure shows the B-spline basis functions used in this work. The thick line highlights the basis function of knot 6.

of a rich and growing literature on shrinkage estimation in macroeconometrics (Ingram & Whiteman, 1994; Del-Negro & Schorfheide, 2004; Hansen, 2016a).

The paper is structured as follows. Section II introduces SLP. Section III contains the simulation study. Section IV presents the empirical application. Concluding remarks follow in section V.

II. Methodology

A. Smooth Local Projections

Let y_t , x_t , and w_{it} for i from 1 to p be stationary time series observed from $t = 1$ to T . Note that the set of variables w_{it} may include lagged values of y_t and x_t . We are interested in the estimation of the dynamic multiplier of y_{t+h} with respect to a change in x_t for h ranging from H_{\min} to H_{\max} , keeping all other variables constant. Typically, H_{\min} is set to either 0 or 1. Also, we define H as $H = H_{\max} - H_{\min}$.

Jordá (2005) proposes to recover the multiplier from the set of regression coefficients $\beta_{(h)}$ associated with the following set of h -step-ahead predictive regressions,

$$y_{t+h} = \alpha_{(h)} + \beta_{(h)}x_t + \sum_{i=1}^p \gamma_{i(h)}w_{it} + u_{(h)t+h}, \quad (1)$$

where $u_{(h)t+h}$ is a prediction error term with $\text{Var}(u_{(h)t+h}) = \sigma_{(h)}^2$.² The set of regressions in equation (1) is named local projections (LP), and Jordá (2005) proposes to recover the dynamic multiplier $\beta_{(h)}$ by running $H + 1$ least squares regressions. LP can be seen as an offshoot of the classic distributed lag (DL) model literature. In particular, if x_t is zero mean or serially uncorrelated and the set of controls w_t is empty, then the estimator of the dynamic multiplier obtained by LP is asymptotically equivalent to the one obtained from an unrestricted DL model.

The dynamic multiplier $\beta_{(h)}$ in equation (1) is expensively parameterized, and in such cases least square estimation may

suffer from excessive variability. In this work, we propose an LP estimation methodology based on penalized B-splines (Eilers & Marx, 1996) that tackles this issue. We call our methodology smooth local projections (SLP). We begin by approximating the $\beta_{(h)}$ coefficient using a linear B-splines basis function expansion in the forecast horizon h , that is,

$$\beta_{(h)} \approx \sum_{k=1}^K b_k B_k(h), \quad (2)$$

where $B_k : \mathbb{R} \rightarrow \mathbb{R}$ for $k = 1, \dots, K$ is a set of B-spline basis functions and b_k for $k = 1, \dots, K$ is a set of scalar parameters. As is customary, in equation (2), we use a relatively large value of K to ensure that the bias due to the approximation step is sufficiently small. We proceed analogously with the $\alpha_{(h)}$ and $\gamma_{i(h)}$ terms so that equation (1) can be approximated as

$$y_{t+h} \approx \sum_{k=1}^K a_k B_k(h) + \sum_{k=1}^K b_k B_k(h)x_t + \sum_{i=1}^p \sum_{k=1}^K c_{ik} B_k(h)w_{it} + u_{(h)t+h}. \quad (3)$$

Note that one may also choose to apply the B-splines basis approximation to a subset of the coefficients of equation (1) rather than all of them.

Figure 1 shows the set of B-splines basis functions used throughout this work. B-splines are a basis of hump-shaped functions indexed by a set of knots. A B-spline basis function is made up of $q + 1$ polynomial pieces of order q . The polynomial pieces join on a set of $q + 2$ inner knots and are calibrated in a way such that derivatives up to the order $q - 1$ are continuous at the inner knots. The B-spline's basis function is nonzero over the domain spanned by the $q + 2$ inner knots and zero elsewhere. The left-most inner knot is used to index the B-spline basis function and the order of the polynomial pieces determines the order of the B-spline basis (i.e., if the polynomial pieces are order q , the B-spline basis is said to be of order q). For illustration purposes, figure 1 highlights the B-spline's basis of knot 6, together with the inner knots

²Several important earlier contributions have put forward this idea, among them, Dufour and Renault (1998) and Cochrane and Piazzesi (2002). See also Al Sadoon (2014) for more recent insights.

used to construct this function. In this work, we use a cubic B-splines basis with equidistant knots ranging from $H_{\min} - 2$ to $H_{\max} - 1$ with unitary increments.³

An appealing feature of the approximating model in equation (3) is that it retains linearity with respect to the parameters and can be represented as a linear regression. Let \mathcal{Y}_t for $t = 1, \dots, T - H_{\min}$ be defined as the vector $(y_{t+H_{\min}}, \dots, y_{\min(T, t+H_{\max})})'$ and let d_t denote its size.⁴ Let $\mathcal{X}_{\beta t}$ be defined as a $d_t \times K$ matrix the (h, k) th element of which is $B_k(h)x_t$. Let $\mathcal{X}_{\alpha t}$, $\mathcal{X}_{\gamma t}$ for $i = 1, \dots, p$ be defined analogous to $\mathcal{X}_{\beta t}$. Let \mathcal{X}_t be the matrix obtained by stacking horizontally the matrices $\mathcal{X}_{\alpha t}$, $\mathcal{X}_{\beta t}$, and $\mathcal{X}_{\gamma t}$ $i = 1, \dots, p$, that is, $\mathcal{X}_t = (\mathcal{X}_{\alpha t} \mathcal{X}_{\beta t} \mathcal{X}_{\gamma t} \dots \mathcal{X}_{\gamma p t})$. Denote by θ the vector of B-splines parameters $(a_1, \dots, a_K, b_1, \dots, b_K, c_{11}, \dots, c_{1K}, \dots, c_{p1}, \dots, c_{pK})'$. Then equation (3) can be compactly represented as a linear model,

$$\mathcal{Y}_t = \mathcal{X}_t \theta + \mathcal{U}_t, \quad (4)$$

where \mathcal{U}_t denotes the $d_t \times 1$ prediction error vector term of the regression. Last, we denote by \mathcal{Y} , \mathcal{X} , and \mathcal{U} the vertically stacked versions of, respectively, \mathcal{Y}_t , \mathcal{X}_t , and \mathcal{U}_t .

We propose to estimate SLP by generalized ridge estimation, that is,

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \{ \|\mathcal{Y} - \mathcal{X}\theta\|^2 + \lambda \theta' \mathbf{P} \theta \} \\ &= (\mathcal{X}' \mathcal{X} + \lambda \mathbf{P})^{-1} \mathcal{X}' \mathcal{Y}, \end{aligned} \quad (5)$$

where λ is a positive shrinkage parameter determining the amount of shrinkage and \mathbf{P} is a symmetric positive semidefinite penalty matrix.⁵ The shrinkage coefficient λ determines the bias/variance trade-off of the estimator: When λ is 0, the estimator coincides with the least square estimator (0 bias but potentially large variance), whereas when λ is large, the estimator is biased but has smaller variance than the least squares estimator. We select λ using k -fold cross-validation (Racine, 1997).

B-splines are attractive in our context because by using a particular family of penalty matrices \mathbf{P} , we can shrink the estimated IR toward a polynomial of a given order instead of shrinking toward 0 like typical shrinkage estimators. To illustrate this feature, consider a constrained version of the model in equation (1) with no intercept and no control variables where the multiplier $\beta_{(h)}$ is the only unknown. As Eilers and Marx (1996) noted, the derivatives of order r of $\sum_{k=1}^K b_k B_k(h)$ with respect to the horizon h can be expressed as linear combinations of the finite differences of order r of adjacent B-spline

coefficients b_k . By considering a penalty matrix of the form $\mathbf{P} = \mathbf{D}'_r \mathbf{D}_r$ with \mathbf{D}_r the matrix representation of the r th difference operator Δ^r ,⁶ one can shrink the estimated IR toward a polynomial of order $r - 1$. Penalizing the r th difference of adjacent B-spline coefficients can be written as

$$\lambda \sum_{k=r+1}^K (\Delta^r b_k)^2 = \lambda (\mathbf{D}_r b)' (\mathbf{D}_r b) = \lambda b' \mathbf{D}'_r \mathbf{D}_r b. \quad (6)$$

A large value of λ will shrink the r th derivative of the estimated multiplier—the polynomial $\sum_{k=1}^K b_k B_k(h)$ —toward 0, which in turn will shrink the estimated multiplier toward a polynomial of order $r - 1$. It is straightforward to generalize this estimation strategy for all coefficients in equation (3). In order to shrink the B-spline approximation of all the $\alpha_{(h)}$, $\beta_{(h)}$, \dots , $\gamma_{p(h)}$ coefficients to, respectively, polynomials of order $r_1 - 1$, $r_2 - 1$, \dots , $r_{2+p} - 1$, the penalty matrix \mathbf{P} can be set to a block diagonal matrix with matrices $\mathbf{D}'_{r_1} \mathbf{D}_{r_1}$, $\mathbf{D}'_{r_2} \mathbf{D}_{r_2}$, \dots , $\mathbf{D}'_{r_{2+p}} \mathbf{D}_{r_{2+p}}$ on the diagonal.⁷

A number of comments on our proposed methodology are in order. Continuing the parallel with the DL literature, consider the case when x_t is mean zero or serially uncorrelated and the set of controls w_t is empty. When the degree of shrinkage is negligible, the SLP estimator of the dynamic multiplier is asymptotically equivalent to the one produced by the unrestricted DL model and standard LP. When the degree of shrinkage is large, the SLP estimator is asymptotically equivalent to the one produced by the polynomial DL model of Almon (1965). By appropriately choosing the amount of penalization, SLP may achieve an optimal balance between these two extremes. SLP can be seen as a modern version of Shiller's (1973) smoothness prior, which was introduced to find a suitable compromise between the unrestricted DL model and Almon's polynomial DL model.⁸ The appealing feature of our framework is that it retains linearity with respect to the parameters and closed-form estimators are readily available.

While we do not derive formal results on the MSE dominance of the SLP estimator over the standard LP, it is important to give some insight into the limitations of shrinkage estimation. The discussion draws largely on recent results established in Hansen (2016b), where the maximum likelihood estimator (MLE) is compared with a class of shrinkage estimators. First, shrinkage may improve on the average MSE across several parameters, but it will rarely uniformly

⁶The \mathbf{D}_r matrix is the matrix such that for a vector x we have that $\Delta^r x_i = [\mathbf{D}_r x]_i$.

⁷Note that further shape constraints can be also implemented. Notably, for stationary series, one may additionally impose that $\beta_{(h)}$ is close to 0 at large enough horizons. This can be easily implemented by shrinking b_k toward 0 (instead of its r th difference) for k large enough.

⁸In fact, our shrinkage estimation approach has a Bayesian interpretation. The sum of squared residuals term in equation (5) can be interpreted as the log likelihood, whereas the penalty term can be thought of as the log density of a Gaussian prior. Thus, the SLP estimator can be thought of as the maximizer of the posterior of the model parameters. Note that from this perspective, we can interpret the penalty matrix \mathbf{P} as a shape prior.

³See Boor (1978) for a textbook presentation of B-splines and Eilers and Marx (1996) for a concise summary on their properties.

⁴Note that the size d_t of the \mathcal{Y}_t vector is not fixed, and it ranges from 1 (for $t = T - H_{\min}$) to $H + 1$ (assuming $T > H$).

⁵Note that the errors of equation (4) are overlapping multistep forecast errors that typically exhibit substantial serial correlation. A GLS-type shrinkage estimator may improve the MSE performance, but we leave this for future research.

improve on the MSE for a single parameter. Second, shrinkage works best when the individual parameter estimators are nearly uncorrelated, since the scope for variance reduction through smoothing is smaller when estimators are highly correlated, as could be the case in some applications with persistent data. Finally, even under ideal conditions, somewhat surprising and subtle conditions are often required for shrinkage estimators to MSE dominate the MLE, as exemplified by the famous James-Stein $n \geq 3$ condition.

While this paper focuses on using SLP for point estimation of $\beta_{(h)}$, confidence intervals are a natural object of interest in empirical applications. Confidence bands for nonparametric estimators like the one here can be challenging to construct, and the required asymptotic theory has only partially been developed (see also Andrews, 1991; Huang, 2003; Chen, 2007). The theory gets more complicated in case of a data-dependent choice of the smoothing parameter. Here we use the following heuristic procedure. We estimate the asymptotic variance of $\hat{\theta}$ using the Newey-West estimator,

$$\widehat{V}(\hat{\theta}) = T \left[\sum_{t=1}^{T-H_{\min}} \mathcal{X}'_t \mathcal{X}_t + \lambda \mathbf{P} \right]^{-1} \left[\hat{\Gamma}_0 + \sum_{l=1}^L w_l (\hat{\Gamma}_l + \hat{\Gamma}'_l) \right] \\ \times \left[\sum_{t=1}^{T-H_{\min}} \mathcal{X}'_t \mathcal{X}_t + \lambda \mathbf{P} \right]^{-1},$$

where $w_l = 1 - l/(1+L)$ and $\hat{\Gamma}_l = \frac{1}{T} \sum_{t=1}^{T-H_{\min}} \mathcal{X}'_t \hat{\mathcal{U}}_t \hat{\mathcal{U}}'_{t-l} \mathcal{X}'_{t-l}$ with $\hat{\mathcal{U}}_t$ denoting regression residuals. In our empirical implementation, we set L to H and λ to 0.1 times the degree of shrinkage determined by k -fold cross-validation.⁹ The pointwise $1 - p$ confidence interval for $\beta_{(h)}$ is then constructed as $B(h)' \hat{b} \pm z_{1-p/2} \sqrt{B(h)' \widehat{V}(\hat{b}) B(h)}$, where $z_{1-p/2}$ denotes the $1 - p/2$ quantile of a standard normal, $B(h) = (B_1(h), \dots, B_K(h))'$, and \hat{b} and $\widehat{V}(\hat{b})$ denote the subvector and submatrix of, respectively, $\hat{\theta}$ and $\widehat{V}(\hat{\theta})$ relative to the b parameter.

B. Estimating Structural Impulse Responses

Carrying out inference on the dynamic effects of structural shocks is key in macroeconomics. In this section, we consider y_t as an endogenous variable in a macroeconomic system of interest, and we are interested in the estimation of its response to a structural shock ε_t .¹⁰ The structural IR of y_t to the shock ε_t is defined as

⁹One of the challenges in the construction of confidence intervals in this context lies in the fact that typically the distribution of shrinkage estimators has a nonnegligible bias, which is a function of the shrinking parameter. Constructing the confidence interval using an undersmoothed estimator of θ reduces the extent of such bias. See also Härdle (1990).

¹⁰We follow the definition of Ramey (2016) and define a structural shock ε_t as a variable (a) that is exogenous with respect to the other current lagged endogenous variables in the system, (b) is uncorrelated with other exogenous shocks, and (c) represents either unanticipated movements in exogenous

$$\text{IR}(h, \delta) = E(y_{t+h} | \varepsilon_t = \delta) - E(y_{t+h} | \varepsilon_t = 0),$$

for $h = H_{\min}$ to H_{\max} . Typically, δ is set to 1 standard deviation of the shock ε_t . Throughout, H_{\min} is 0 unless specified otherwise. In this section, we illustrate how (S)LP may be used to recover structural IRs when the structural shock is observed or can be recovered through controls, or can be recovered through an instrument.

Identification through controls. In the identification through controls case, the IR can be estimated by running regression (1) with the appropriate set of controls (see Angrist, Jordá, & Kuersteiner, 2018; Jordá & Taylor, 2016). More precisely, if the shock is observed, then the IR can be estimated simply by running equation (1), setting x_t equal to the structural shock (with w_t empty). If the shock is identified as the residual of the regression of an endogenous variable on a set of control variables, then the IR can be estimated by running equation (1), setting x_t equal to the endogenous variable and w_t equal to the set of controls. Note that additional regressors may be included in equation (1) to “mop up” the residual variance. In both settings, the coefficient $\beta_{(h)}$ captures the causal effect of the structural shock and the IR(h, δ) is given by $\beta_{(h)} \delta$, which can be estimated as $\widehat{\text{IR}}(h, \delta) = \hat{\beta}_{(h)} \delta$.

Interestingly, the recursive identification scheme put forward by Sims (1980) can be seen as a special case of identification through controls. Sims (1980) proposes timing restrictions between the exogenous shocks of the VAR to disentangle the causal chain of events and identify the structural shocks of interest. In the LP setting, such timing restrictions can be imposed with a specific choice of control variables and H_{\min} . Although known at least since Shapiro and Watson (1988), this point has been relatively underappreciated among LP practitioners.

As an illustration, consider a system comprising output gdp_t , inflation π_t , and the Fed funds rates ffr_t . The objective is to estimate the IR of output to a monetary shock to the Fed funds rate. Assuming that the system evolves according to a VAR of order 1 and that the monetary shocks do not affect the other variables on impact, one can recover the IR of output from the LP by setting $y_{t+h} = \text{gdp}_{t+h}$, $x_t = \text{ffr}_t$, and $w'_t = (\text{gdp}_t, \pi_t, \text{gdp}_{t-1}, \pi_{t-1}, \text{ffr}_{t-1})'$ for h ranging from $H_{\min} = 1$ until H_{\max} . Intuitively, we achieve identification by controlling for the contemporaneous values of variables ordered before the shock of interest (in this case, output and inflation).¹¹

Identification through instruments. Even when a shock or an appropriate set of control variables is not available, it may still be possible to recover the IR through an instrument by

variables or news about future movements in exogenous variables (see also Blanchard & Watson, 1986; Bernanke, 1986, and Stock & Watson, 2016).

¹¹This strategy effectively amounts to identifying monetary shocks from the residuals of a Taylor rule with output growth and inflation (and their lags).

running a two-stages least squares regression (Stock & Watson, 2018; Plagborg-Møller & Wolf, 2018). For instance, we may have that the macroeconometrician observes a noisy measurement of the structural shock m_t defined as $\varepsilon_t + e_t$, where e_t is a conditionally unpredictable measurement error.¹² In these cases, an instrument can allow recovering the effect of interest. We define an instrument z_t to be a time-series satisfying $\text{corr}(\varepsilon_t, z_t) \neq 0$ and $\text{corr}(e_t, z_t) = 0$ so that z_t is relevant and exogenous. Then the IR can be estimated using two-stages least squares estimation. More precisely, in a first-stage regression, we regress m_t on the instrument z_t . In the second-stage regression, we run equation (1), setting x_t equal to the fitted value of the first-stage regression. Again, the $\beta_{(h)}$ coefficient captures the structural effect of the shock.

To illustrate our instrumental approach, we use our previous monetary example and consider the series of monetary shocks narratively identified by Romer and Romer (2004). A reasonable assumption may be to posit that the Romer and Romer (2004) shocks are a proxy for the true monetary shocks rather than an exact measure, in that they are correlated with the true monetary shocks and uncorrelated with other structural shocks. In that case, the Romer and Romer (2004) shock series satisfies the instrumental variable conditions, and we can recover IR to monetary shocks from (S)LP and two-stages least squares where ffr_t is instrumented with the Romer and Romer (2004) shocks series z_t . Specifically, in the first stage, we regress ffr_t on z_t , and in the second stage, we estimate SLP with $y_{t+h} = \text{gdp}_{t+h}$, $x_t = \hat{\text{ffr}}_t$ where $\hat{\text{ffr}}_t$ is the fitted value of the first-stage regression.

III. Simulation Study

In this section, we carry out a simulation study to benchmark the performance of IR estimation based on SLP against LP and VARs. We consider a system with GDP growth gdp_t , PCE inflation π_t , and the Fed funds rate ffr_t . The system is generated as

$$\begin{aligned} \text{gdp}_t &= \sum_{h=0}^{20} \beta_{11(h)} \varepsilon_{t-20+h}^{\text{gdp}} + \sum_{h=1}^{20} \beta_{12(h)} \varepsilon_{t-20+h}^{\pi} \\ &\quad + \sum_{h=1}^{20} \beta_{13(h)} \varepsilon_{t-20+h}^{\text{ffr}}, \\ \pi_t &= \sum_{h=0}^{20} \beta_{21(h)} \varepsilon_{t-20+h}^{\text{gdp}} + \sum_{h=0}^{20} \beta_{22(h)} \varepsilon_{t-20+h}^{\pi} \\ &\quad + \sum_{h=1}^{20} \beta_{23(h)} \varepsilon_{t-20+h}^{\text{ffr}}, \end{aligned}$$

¹²Let us emphasize that we assume e_t to be unpredictable given past information. If the measurement error is correlated with past shocks, then identification becomes more involved.

$$\begin{aligned} \text{ffr}_t &= \sum_{h=0}^{20} \beta_{31(h)} \varepsilon_{t-20+h}^{\text{gdp}} + \sum_{h=0}^{20} \beta_{32(h)} \varepsilon_{t-20+h}^{\pi} \\ &\quad + \sum_{h=0}^{20} \beta_{33(h)} \varepsilon_{t-20+h}^{\text{ffr}}, \end{aligned} \quad (7)$$

where $\varepsilon_t^{\text{gdp}}$, ε_t^{π} , and $\varepsilon_t^{\text{ffr}}$ are i.i.d. structural normal shocks with mean zero and variances equal to, respectively, σ_{gdp}^2 , σ_{pi}^2 , and σ_{ffr}^2 . Notice that standard recursive timing restrictions are imposed on the contemporaneous impact of the shocks in the system. These restrictions allow identifying the full set of structural IRs using standard methods (see section IIB). The exercise focuses on the estimation of the IR of gdp_t to a monetary shock $\varepsilon_t^{\text{ffr}}$.

In order to entail realistic dynamics for the simulations, the $\beta_{ij(h)}$ parameters of equation (7) are based on the coefficients of the nine structural IRs estimated with LP over 1959Q1–2007Q4. Specifically, we identify the IRs of the structural shocks in equation (7) through controls (see section IIB) by including in the LP regression the appropriate subset of contemporaneous series as well as four lags of all variables in the system. For example, the IRs associated with inflation shocks ε_t^{π} are identified by setting $x_t = \pi_t$ and $w_t' = (\text{gdp}_t, \text{gdp}_{t-1}, \pi_{t-1}, \text{ffr}_{t-1}, \dots, \text{gdp}_{t-4}, \pi_{t-4}, \text{ffr}_{t-4})'$.

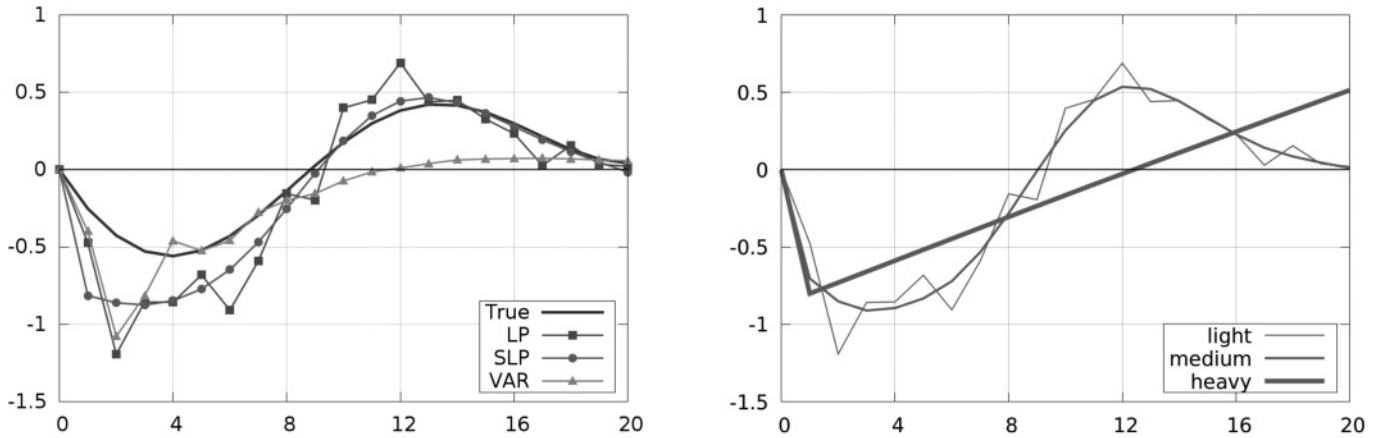
To assess how the performance of SLP varies with the degree of smoothness of the IR, we consider four sets of simulation in which the multiplier of interest— $\beta_{13(h)}$, the response of GDP growth to a fed funds rate shock—is made increasingly jagged, whereas the other multipliers are kept unchanged at their LP point estimates. To simulate plausible degrees of noise in the IR of interest, we proceed as follows. We construct a “smooth IR” by smoothing the $\beta_{13(h)}$ estimates from LP,¹³ to which we add Gaussian noise at each horizon. As a benchmark, a baseline value for the noise variance is the variance of the difference between the LP IR estimate and its smoothed counterpart. In the first DGP, labeled A, the IR of GDP growth to a Fed funds rate shock is the smooth (noiseless) IR. In DGP (B), the IR is the smooth IR used in DGP (A) plus Gaussian noise with a standard deviation set at one-half its benchmark level. In DGP (C), the IR is set equal to its LP point estimate, so that the noise variance is at its benchmark level. In DGP (D), the IR is the smooth IR used in DGP (A) plus Gaussian noise with a standard deviation set

¹³The smooth IR is obtained by regressing the LP IR estimates on a sine/cosine basis,

$$\begin{aligned} \hat{\beta}_{13(h)} &= c_1 \sin\left(\frac{2\pi}{H}h\right) + c_2 \cos\left(\frac{2\pi}{H}h\right) + c_3 \sin\left(\frac{2\pi}{H}2h\right) \\ &\quad + c_4 \cos\left(\frac{2\pi}{H}2h\right) + e_h, \end{aligned}$$

and then using the fitted values of the regression as the smooth IR. We use a different smoothing method than B-splines in order not to mechanically bias results in our favor.

FIGURE 2.—SLP AND THE DEGREE OF SHRINKAGE



The figure displays the estimated IR for one replication of the simulation study for parameter setting (B). The panels show the estimated IR of LP (square) and SLP (circle), as well as the population IR (solid line) for GDP growth (left panel). The right panel displays the estimated IRs obtained from SLP using different degrees of shrinkage, with thicker lines denoting estimates obtained using a higher degree of penalization.

at twice its benchmark level. This set of simulations allows us to study the performance of SLP for IRs with different degrees of smoothness, from a smoother IR–DGP (A)– to a noisier IR–DGP (D)–.

We estimate the IR of gdp_t to a monetary shock with SLP using timing restrictions consistent with our DGP. A number of details on the implementation of the SLP estimator used in this study are in order. First, we use smooth regularization only on the coefficients associated with the IR of interest, and we do not smooth the coefficients of the control variables.¹⁴ Regarding the choice of the penalty matrix, we opt for a naive approach and shrink toward a line (i.e., $r = 2$), which is roughly consistent with the IR estimated by the standard LP. Finally, the shrinkage parameter λ is chosen by five-fold cross-validation. For comparison purposes, we also report the estimation results of the Oracle SLP estimator, that is, the SLP estimator estimated using the shrinkage parameter λ that minimizes the MSE of the IR estimator. The Oracle shrinkage level is determined by simulation. We benchmark our methodology against standard LP estimated by least squares, VAR(4), VAR(12), and a VAR with order chosen via the AIC.

We replicate our simulation exercise for our six-parameter setting using a sample size equal to 50, 100, 200, and 400. The simulation is replicated 1,000 times for each parameter setting and sample size. The performance of each IR estimator is measured by its integrated MSE defined as $E[\sum_{h=H_{min}}^{H_{max}} (\widehat{IR}(h, \delta) - IR(h, \delta))^2]$, which is approximated using the Monte Carlo average across replications.

We use one replication for illustration purposes. The left panel of figure 2 shows the IR estimates based on SLP (based on cross-validation), LP, and VAR(4). Note that despite the

¹⁴This allows us to more easily compare the LP and SLP estimators but makes the exercise more disadvantageous for our SLP methodology as further efficiency gains may be attained by using regularization more extensively.

TABLE 1.—IR ESTIMATION COMPARISON

	T	SLP			VAR		
		LP	Ridge/ CV	Ridge/ Oracle	$P = 4$	$P = 12$	AIC
A (smoother)	50	6.46	59.77	65.25	65.09	-13.93	-14.14
	100	3.15	55.37	61.66	44.18	25.54	25.46
	200	1.64	51.45	58.74	7.97	32.56	32.52
	400	0.92	43.22	56.04	-49.55	33.08	33.09
B	50	6.73	56.80	62.72	63.13	-16.10	-16.23
	100	3.17	48.53	56.02	39.36	24.33	24.20
	200	1.73	38.04	46.76	-0.07	28.54	28.47
	400	0.97	28.79	41.14	-62.44	24.31	24.32
C	50	6.79	43.50	49.04	61.05	-16.91	-17.07
	100	3.32	23.07	28.33	36.49	20.07	19.97
	200	1.73	4.26	15.70	-6.86	22.45	22.38
	400	0.93	4.32	10.10	-86.62	12.42	12.46
D (rougher)	50	9.33	29.13	33.52	29.87	-16.52	-16.71
	100	4.77	6.94	13.36	-26.80	10.81	10.77
	200	2.64	1.90	5.48	-120.98	1.03	1.08
	400	1.63	-2.88	0.86	-255.18	-20.81	-20.82

The first column reports the MSE of the IR of GDP growth to a monetary policy shocks estimated via LP (based on least squares), while the remaining columns report the percentage improvement (relative to LP) from SLP (based on cross-validated generalized ridge and Oracle generalized ridge) and VAR (using a lag length of 4, 12, and the one determined by the AIC). A positive entry denotes improvement over LP.

population IR being smooth, the LP delivers IR estimates that are quite rough, a well-known feature of LP (Ramey, 2016). We can see that SLP essentially smooths the LP, and in this particular replication, it delivers a more precise estimate of the IR. The right panel of figure 2 shows how the SLP IR estimates change depending on the degree of shrinkage λ . When λ is small, the SLP estimate is practically indistinguishable from the regular LP estimate, but as λ increases, the estimated IR becomes progressively smoother and closer to the target polynomial implied by the choice of the penalty matrix (in this case, a line).

Table 1 reports summary results for the simulation study. The first column contains the MSE of the standard LP, whereas the remaining columns contain the percentage improvements of the alternative estimation methods. Standard LP is typically outperformed by the majority of alternative IR

TABLE 2.—CONFIDENCE INTERVAL LENGTH AND COVERAGE

T			Horizon								
			2	4	6	8	10	12	14	16	18
50	LP	Length	1.106	1.114	1.136	1.121	1.154	1.180	1.178	1.156	1.162
		Coverage	0.790	0.761	0.762	0.761	0.790	0.788	0.785	0.717	0.760
	SLP	Length	0.807	0.822	0.837	0.836	0.864	0.875	0.876	0.884	0.907
100	LP	Coverage	0.588	0.706	0.789	0.718	0.767	0.751	0.752	0.684	0.706
		Length	0.911	0.939	0.947	0.965	0.974	0.982	0.972	0.975	0.993
	SLP	Coverage	0.860	0.806	0.847	0.820	0.837	0.857	0.845	0.803	0.844
200	LP	Length	0.722	0.747	0.760	0.768	0.785	0.802	0.796	0.806	0.831
		Coverage	0.629	0.750	0.847	0.819	0.828	0.824	0.828	0.771	0.785
	SLP	Length	0.735	0.748	0.751	0.768	0.784	0.780	0.773	0.782	0.800
400	LP	Coverage	0.895	0.878	0.878	0.902	0.868	0.886	0.870	0.822	0.881
		Length	0.654	0.669	0.673	0.690	0.706	0.708	0.707	0.713	0.747
	SLP	Coverage	0.743	0.840	0.879	0.871	0.853	0.872	0.869	0.814	0.854
400	LP	Length	0.565	0.575	0.576	0.586	0.598	0.595	0.592	0.595	0.606
		Coverage	0.914	0.915	0.869	0.913	0.896	0.902	0.891	0.840	0.887
	SLP	Length	0.538	0.548	0.549	0.562	0.574	0.573	0.571	0.575	0.593
		Coverage	0.865	0.906	0.875	0.888	0.892	0.899	0.885	0.821	0.886

The table reports average length and average coverage of the 90% confidence intervals for LP and SLP for sample size $T = 50, 100, 200,$ and 400 and horizons h ranging from 2 to 18 (in increments of 2).

estimation methods. The gains of SLP can be quite substantial, especially when the sample size is small. In addition, and not surprisingly, gains are larger when the population IR is smoother (i.e., for DGPs (A) and (B)). Comparing the performance of the Oracle versus the cross-validated SLP, we see that there are no large differences between using the optimal λ and selecting a λ from cross-validation, indicating that for the class of DGPs considered in this study, cross-validation performs satisfactorily. While VAR-based IR estimators can perform well at times, their performances are sensitive to the choice of the number of lags. The VAR(4) does remarkably well when the sample size is small, but the gains relative to SLP deteriorate when the sample size increases. On the contrary, the VAR(12) and VAR with AIC lag selection perform better when the sample size is larger.¹⁵

Finally, we investigate the properties of the confidence intervals procedure we propose. We simulate DGP (A) and we construct the LP and SLP 90% confidence intervals. The LP confidence intervals are constructed using Newey-West standard errors with a number of lags equal to H , whereas the SLP confidence intervals are based on the procedure previously described. Table 2 reports the average length of the confidence interval, as well as the coverage of the interval over 1,000 replications. The simulations show that the LP and SLP confidence interval procedures have similar performances. While the SLP confidence intervals are narrower, they also have slightly smaller coverage. There can be pronounced size distortions for smaller samples, but they become less severe as the sample size increases.

IV. Empirical Illustration

In this section we use our proposed methodology to study the effects of monetary shocks on output, which have been the

¹⁵On average the AIC tends to select large VAR orders across the different parameter settings.

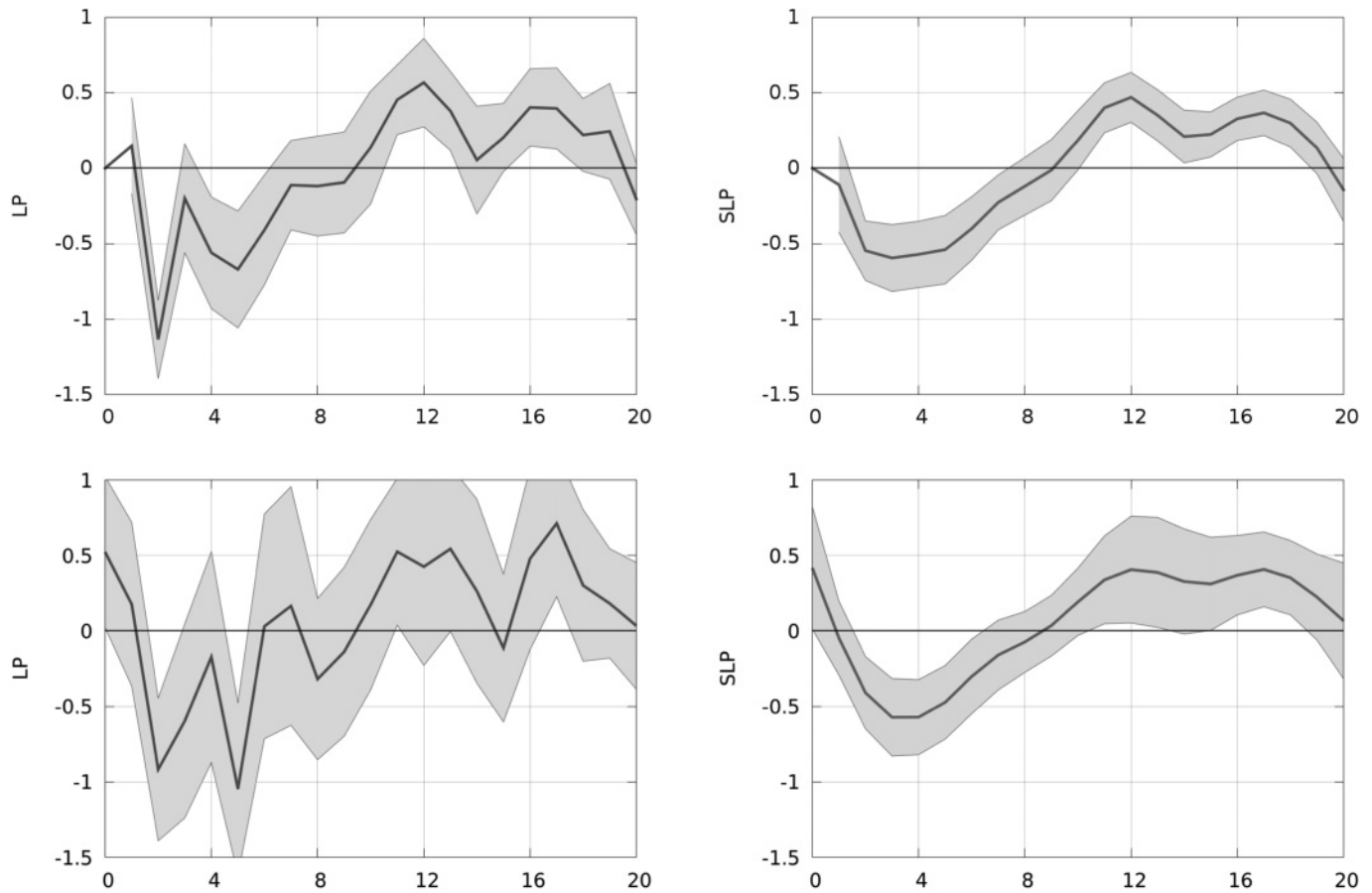
subject of extensive research (see Ramey, 2016, for a review). Here we apply our SLP approach using identification with timing restrictions and IV. In the timing restrictions case, we assume that we can identify the IR of GDP growth to a monetary shock from an SLP of GDP growth on the Fed funds rate using as controls the contemporaneous value of GDP growth and inflation as well as four lags of GDP growth, inflation, and the Fed funds rate. In the IV case, we use the Romer and Romer monetary shocks series as instrument for movements in the Fed funds rate (Romer & Romer, 2004; Coibion, Gorodnichenko, & Silvia, 2012). As controls, we include four lags of GDP growth, inflation, and the Fed funds rate. The sample spans 1966-Q1 to 2007-Q4.

Figure 3 plots the IRs of GDP growth and inflation to a 1 standard deviation monetary shock. The left panel plots the impulse responses obtained from LPs, while the right panel plots the IRs obtained from SLP. Following a contractionary shock, GDP growth declines, as previously found in numerous studies. However, the IRs obtained by regular LP can be erratic, with sometimes sharp fluctuations from quarter to quarter. This makes the interpretation of certain features of the IR difficult, since it is not clear whether these movements are real features of the IR or just artifacts of noisy measurements (e.g., Ramey, 2012). In contrast, thanks to smoothing, the SLP IRs are easier to interpret.

V. Conclusion

This paper proposes a novel IR estimation approach based on penalized B-splines called smooth local projections (SLP). The SLP approach preserves the flexibility of standard LP but can substantially increase precision. Moreover, SLP estimation boils down to standard ridge regression. A simulation study is used to illustrate the performance of SLP for IR estimation, and we find that SLP can deliver substantial improvements over LP. As with LP, SLP can be easily used with common identification schemes to directly estimate structural

FIGURE 3.—IR OF GDP GROWTH TO A MONETARY SHOCK



The figure displays the IR of GDP growth to a monetary shock identified using timing restrictions (top panels) and instrumental variables (bottom panels) estimated using LP (left panels) and SLP (right panels). The shaded area denotes the 90% confidence interval.

IRs. We illustrate our approach by studying the effects of monetary shocks with different identification schemes.

REFERENCES

Al Sadoon, M., “Geometric and Long Run Aspects of Granger Causality,” *Journal of Econometrics* 178 (2014), 558–568.

Almon, S., “The Distributed Lag between Capital Appropriations and Net Expenditures,” *Econometrica* 33 (1965), 178–19.

Andrews, D. W. K., “Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Models,” *Econometrica* 59 (1991), 307–345.

Angrist, J., O. Jordá, and G. Kuersteiner, “Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited,” *Journal of Business and Economic Statistics* 36:3 (2018), 371–387.

Auerbach, A. J., and Y. Gorodnichenko, “Fiscal Multipliers in Recession and Expansion” (pp. 63–98), in *Fiscal Policy after the Financial Crisis* (Cambridge, MA: NBER, 2012).

Barnichon, R., and C. Matthes, “Functional Approximation of Impulse Responses,” *Journal of Monetary Economics* 99 (2018), 41–55.

Bernanke, B. S., “Alternative Explanations of the Money-Income Correlation,” *Carnegie-Rochester Conference Series on Public Policy* 25:1 (1986), 49–99.

Blanchard, O. J., and M. W. Watson, “Are Business Cycles All Alike?” (pp. 123–180), in Robert, J. Gordon, ed., *The American Business Cycle: Continuity and Change* (Cambridge, MA: NBER, 1986).

Boor, C. D., *A Practical Guide to Splines* (Berlin: Springer, 1978).

Chen, X., “Large Sample Sieve Estimation of Semi-Nonparametric Models” (pp. 1–2), in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics*, vol. 6B (Amsterdam North-Holland, 2007).

Cochrane, J. H., and M. Piazzesi, “The Fed and Interest Rates: A High-Frequency Identification,” *American Economic Review* 92:2 (2002), 90–95.

Coibion, O. Y., L. K. Gorodnichenko, and J. Silvia, “Innocent Bystanders? Monetary Policy and Inequality in the US,” NBER working papers 18170 (2012). Research, Inc.

Del-Negro, M., and F. Schorfheide, “Priors from General Equilibrium Models for VARs,” *International Economic Review* 45:2 (2004), 643–673.

Dufour, J.-M., and E. Renault, “Short Run and Long Run Causality in Time Series: Theory,” *Econometrica* 66:5 (1998), 1099–1126.

Eilers, Paul H. C., and Brian D. Marx, “Flexible Smoothing with B-Splines and Penalties,” *Statistical Science* 11 (1996), 89–102.

Hansen, B., “Stein Combination Shrinkage for Vector Autoregressions,” University of Wisconsin working paper (2016a).

———, “Efficient Shrinkage in Parametric Models,” *Journal of Econometrics* 190 (2016b), 115–132.

Härdle, W., *Applied Nonparametric Regression* (Berlin: Springer, 1990).

Huang, J., “Local Asymptotics for Polynomial Spline Regression,” *Annals of Statistics* 31 (2003), 1600–1635.

Ingram, B., and C. Whiteman, “Supplanting the Minnesota Prior: Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors,” *Journal of Monetary Economics* 34:3 (1994), 497–510.

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

Jordá, O., and A. M. Taylor, “The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy,” *Economic Journal* 126:590 (2016), 219–255.

Miranda-Agrippino, S., and G. Ricco, “The Transmission of Monetary Policy Shocks,” Bank of England staff working paper 657 (2017).

- Plagborg-Møller, M., "Essays in Macroeconometrics," PhD diss. Harvard University (2016).
- Plagborg-Møller, M., and C. K. Wolf, "Instrumental Variable Identification of Dynamic Variance Decompositions," Princeton University technical report (2018).
- Racine, J., "Feasible Cross-Validatory Model Selection for General Stationary Processes," *Journal of Applied Econometrics* 12:2 (1997), 169–179.
- Ramey, V., "Comment on 'Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment'" (pp. 147–153), in Daron Acemoglu, Jonathan Parker, and Michael Woodford, eds., *NBER Macroeconomics Annual 2012* (Cambridge, MA: NBER, 2012).
- "Macroeconomic Shocks and Their Propagation," H. Uhlig and J. Taylor, eds., *Handbook of Macroeconomics* (Amsterdam: Elsevier, 2016).
- Ramey, V., and S. Zubairy, "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy* 126:2 (2018), 850–901.
- Romer, C. D., and D. H. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review* 94:4 (2004), 1055–1084.
- "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review* 100:3 (2010), 763–801.
- Shapiro, M., and M. Watson, "Sources of Business Cycles Fluctuations" (pp. 111–156), in Stanley Fischer, ed., *Macroeconomics Annual 1988*, vol. 3 (Cambridge, MA: NBER, 1988).
- Shiller, R. J., "A Distributed Lag Estimator Derived from Smoothness Priors," *Econometrica* 41 (1973), 775–788.
- Sims, C., *Distributed Lags: Frontiers of Quantitative Economics II* (Amsterdam: North-Holland, 1974).
- "Macroeconomics and Reality," *Econometrica* 48 (1980), 1–48.
- Stock, J. H., and M. W. Watson, "Factor Models and Structural Vector Autoregressions in Macroeconomics," in H. Uhlig and J. Taylor, eds., *Handbook of Macroeconomics* (Amsterdam: Elsevier, 2016).
- "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal* 128:610 (2018), 917–948.

Copyright of Review of Economics & Statistics is the property of MIT Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.