DISCUSSION PAPER SERIES

DP11726

IMPULSE RESPONSE ESTIMATION BY SMOOTH LOCAL PROJECTIONS

Régis Barnichon and Christian Brownlees

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

IMPULSE RESPONSE ESTIMATION BY SMOOTH LOCAL PROJECTIONS

Régis Barnichon and Christian Brownlees

Discussion Paper DP11726
Published 28 December 2016
Submitted 28 December 2016

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

IMPULSE RESPONSE ESTIMATION BY SMOOTH LOCAL PROJECTIONS

Abstract

Vector Autoregressions (VAR) and Local Projections (LP) are well established methodologies for the estimation of Impulse Responses (IR). These techniques have complementary features: The VAR approach is more efficient when the model is correctly specified whereas the LP approach is less efficient but more robust to model misspecification. We propose a novel IR estimation methodology -- Smooth Local Projections (SLP) -- to strike a balance between these approaches. SLP consists in estimating LP under the assumption that the IR is a smooth function of the forecast horizon. Inference is carried out using semi-parametric techniques based on Penalized B-splines, which are straightforward to implement in practice. SLP preserves the flexibility of standard LP and at the same time can increase precision substantially. A simulation study shows the large gains in IR estimation accuracy of SLP over LP. We show how SLP may be used with common identification schemes such as timing restrictions and instrumental variables to directly recover structural IRs. We illustrate our technique by studying the effects of monetary shocks.

JEL Classification: C14, C32, C53, E47

Keywords: impulse response, local projections, semiparametric estimation

Régis Barnichon - rbarnichon@crei.cat
San Francisco Fed, CREI, Universitat Pompeu Fabra and CEPR

Christian Brownlees - christian.brownlees@upf.edu
Universitat Pompeu Fabra

Impulse Response Estimation By Smooth Local Projections

Regis Barnichon[†] Christian Brownlees[‡]

December 2, 2016

Abstract

Vector Autoregressions (VAR) and Local Projections (LP) are well established methodologies for the estimation of Impulse Responses (IR). These techniques have complementary features: The VAR approach is more efficient when the model is correctly specified whereas the LP approach is less efficient but more robust to model misspecification. We propose a novel IR estimation methodology – Smooth Local Projections (SLP) – to strike a balance between these approaches. SLP consists in estimating LP under the assumption that the IR is a smooth function of the forecast horizon. Inference is carried out using semi-parametric techniques based on Penalized B-splines, which are straightforward to implement in practice. SLP preserves the flexibility of standard LP and at the same time can increase precision substantially. A simulation study shows the large gains in IR estimation accuracy of SLP over LP. We show how SLP may be used with common identification schemes such as timing restrictions and instrumental variables to directly recover structural IRs. We illustrate our technique by studying the effects of monetary shocks.

Keywords: local projections, semiparametric estimation, structural impulse response

JEL: C14, C32, C53, E52

[†] San Francisco Fed, CREI, Universitat Pompeu Fabra, CEPR;
[‡] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE;
e-mail: `christian.brownlees@upf.edu`.
We would like to thank Jordi Gali, Oscar Jorda, Barbara Rossi and seminar participants for helpful comments. Christian Brownlees 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), 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 upon request to the authors.

1 Introduction

Impulse Response (IR) functions are a key tool to summarize 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 the use of an alternative nonparametric IR estimation approach: the Local Projections (LP) of Jordá (2005).[1]

In its basic formulation, the LP approach consists in running a sequence of predictive regressions of a variable of interest on a structural shock for different prediction horizons. The IR is then given by the sequence of regression coefficients of the structural shock. Unlike VARs, the LP method does not impose any underlying dynamics on the variables in the system. This confers a number of advantages: The technique is more robust to misspecification, does not suffer from the curse of dimensionality inherent to VARs, and can more easily accommodate nonlinearities.[2] However, the nonparametric nature of the LP method comes at an efficiency cost: LP is still expensively parametrized and the IR estimator can be excessively noisy (Ramey, 2012, 2016).

In this work we introduce an IR estimation methodology called Smooth Local Projections (SLP) that aims at striking a balance between the VAR and LP approaches. Intuitively, our approach aims at improving the efficiency of LP by imposing that impulse responses are smooth functions of the forecast horizons. Specifically, we work within the LP framework but with the additional assumption that the IR is a smooth function of the prediction horizon. To carry out inference, we rely on a semi-parametric regression technique called penalized B-splines regression (Eilers and Marx, 1996). Our approach

[1] The narrative identification approach was first proposed by Romer and Romer (1989). For recent uses of the narrative identification approach, see Romer and Romer (2004) and Monnet (2014) for monetary shocks, Romer and Romer (2010) for tax shocks, Ramey (2011) and Ramey and Zubairy (2014a) for government spending shocks, Romer and Romer (2014) for shocks to government transfer payments and Romer and Romer (2015) for financial market shocks.

[2] Auerbach and Gorodnichenko (2012), Ramey and Zubairy (2014b) and Jordá and Taylor (2016) use non-linear LPs to estimate the extent of state dependence in the effect of shocks to government spending, while Tenreyro and Thwaites (2016) and Santoro, Petrella, Pfajfar, and Gaffeo (2014) explore the extent of state dependence in the effect of monetary shocks.

consists in (i) approximating the sequence of LP coefficients using a linear combination of B-splines basis functions and (ii) estimating the model using a shrinkage estimator that penalizes non-smooth behavior of the LP coefficients. The penalized B-splines approach has a number of highlights. First, SLP estimation boils down to a linear ridge regression that is straightforward to implement in practice. Second, by appropriately choosing the penalty function of the ridge estimator, the LP coefficients can be shrunk towards a polynomial of a given order (choosen by the user). This is attractive for macroeconomic applications when prior knowledge on the shape of the IR is available.[3] When the degree of penalization imposed in the ridge regression is sufficiently large, SLP reduces to the Almon (1965) distributed lag model, in which the impulse response function is parametrized as a polynomial function.

The SLP approach can be used to recover structural IRs with a number of identification schemes. In this work we show how to recover the causal impact of a structural shock when the shock (i) is observed, (ii) is not observed but can be recovered through control variables, or (iii) is not observed and it can be recovered through instrumental variables. Three applications of SLP are of particular interest for macro-economists. First, we show that the popular recursive identification scheme found in many structural VARs is a particular case of (ii), so that the structural IR can be obtained directly from a (S)LP, provided that the right set of control variables is included. Second, we show that the increasingly popular proxy-SVAR (Stock and Watson, 2008; Mertens and Ravn, 2013) can be seen as a particular case of (iii) and can therefore be easily implemented in the (S)LP framework.[4] Last, we show how the SLP approach can be used to estimate certain classes of nonlinear IRs, and in particular, we show how to identify and estimate state dependent IRs, which have recently been of great interest to macro-economists. While most studies have relied on regime-switching models, which are expensively parametrized

[3]For instance, the impulse response functions to shocks is often theoretically predicted to be monotonic or hump-shaped (e.g. Christiano, Eichenbaum, and Evans, 1999; Walsh, 2010) in the context of monetary policy), and in such case, one may want to shrink the estimated IR towards a second-order polynomial.

[4]In other words, while researchers have relied on VARs to implement identification schemes based on a Cholesky ordering or an external instrument, we show that it is possible to bypass the VAR entirely and thus reduce the risk of model mis-specification.

and can be computationally intensive, SLP offers a flexible semi-parametric alternative that is straightforward to implement.

A simulation study is used to illustrate the performance of SLP for IR estimation. We consider two DGP settings labeled as "rough" and "smooth". In the rough setting IRs can change abruptly across horizons whereas in the smooth case IRs change smoothly. We find that SLP delivers substantial improvements over LP for both DGPs, with especially large gains in the smooth setting. SLP performs better even when the DGP is rough, because LP is so expensively parametrized that shrinkage estimation can produce substantial gains even when the IR is not smooth.

We illustrate our methodology by studying the effects of monetary shocks on GDP growth and inflation. We use two different identification strategies: (i) timing restrictions and (ii) instrumental variables using the Romer and Romer (2004) narrative shocks as an instrument. We use two identification schemes in order to highlight the benefit of being able to implement different identification schemes using the same methodology, something that was not possible before and that could create some degree of confusion when comparing studies.[5] We find that a contractionary shock lowers GDP growth and inflation by similar magnitudes across the two identification methods. Our results also highlights that while LP-based IRs can be erratic, SLP-based IRs are less noisy and thus much easier to interpret. We then explore the state-dependent effects of monetary policy, and find again that SLP delivers a clear message: the effect of a contractionary shock on GDP growth is substantially larger in recessions, i.e., when growth is weak. The response of inflation is also weaker in recessions (the price puzzle is larger), which is consistent, at least qualitatively, with a Keynesian narrative: monetary policy will have a stronger impact on real quantities if prices react less.

Our paper is related to different strands of the literature. First, it contributes to a rapidly growing macroeconomic literature that relies on LPs to estimate structural impulse

[5] As studied in Coibion, Gorodnichenko, and Silvia (2012), this problem was particularly salient in the monetary policy literature. The range of estimates was wide across identification schemes, but the source of the discrepancy was unclear, because the different identification schemes were implemented using different methods. By being able to implement two different identifying schemes using the same methodology, LP provides a framework that enables a easier comparison across identification schemes.

responses. Influential contributions in this literature includes the work of Auerbach and Gorodnichenko (2012), Jordá and Taylor (2016) and Ramey (2016). In terms of methodology, our approach relates to the recent Gaussian Mixture Approximation (GMA) method of Barnichon and Matthes (2016) which consists in approximating IRs with a small number of Gaussian basis functions. While the GMA is a non-linear model, SLP has the advantage of being linear, which preserves the ease of estimation and flexibility of the original LP. Finally, our approach can be cast in the broader literature on semi-parametric and sieve estimation in statistics and econometrics. Among others, White (2006), Chen (2007), Hastie and Friedman (2009) and Hansen (2014) contain general surveys in this area.

The rest of the paper is structured as follows. Section 2 introduces SLP methodology. Section 3 contains the simulation study. Section 4 contains the empirical application. Concluding remarks follow in Section 5.

2 Methodology

2.1 Smooth Local Projections

Let y_t, x_t and $w_{i\,t}$ for i from 1 to p be stationary time series observed from $t = 1$ to T. Note that the set of variables $w_{i\,t}$ 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 0 to H, keeping all other variables constant. There are two main approaches to recover the dynamic multiplier.

In the first approach, the multiplier is obtained from the regression coefficient $\beta_{(h)}$ of the following set of h-step ahead predictive regressions

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

where $u_{(h)\,t+h}$ is a prediction error term with $\mathrm{Var}(u_{(h)\,t+h}) = \sigma^2_{(h)}$. The set of regressions in (1) is named Local Projections (LP) by Jordá (2005) and it may be estimated by running $H + 1$ separate linear regressions. In analogy to the forecasting literature we label this approach "direct", as it obtains the multiplier of interest by specifying a regression model for the dependent variable at the horizon of interest (see Marcellino, Stock, and Watson, 2006).

The alternative approach that can be used to recover the multiplier of y_{t+h} on x_t consists in specifying a dynamic model for the variables in system. In empirical applications this is typically carried out by assuming that the system evolves according to a VAR. For example, let y_t and x_t be generated by a zero-mean bivariate stable VAR

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \mathbf{A} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + u_t,$$

where $\mathbf{A}$ is a 2×2 matrix and u_t is an innovation term. Then, by recursive substitution we have that

$$\begin{bmatrix} y_{t+h} \\ x_{t+h} \end{bmatrix} = \mathbf{A}^h \begin{bmatrix} y_t \\ x_t \end{bmatrix} + u_{(h)\,t+h},$$

where $u_{(h)\,t+h}$ is the h-step ahead innovation. Thus, the dynamic multiplier of y_{t+h} with respect to x_t keeping y_t constant is the (1,2) element of the matrix $[\mathbf{A}^h]$. In analogy to the forecasting literature we label this approach "iterated", as one obtains the multiplier of interest by iterating the dynamic system forward.

The direct and iterated IR estimation approaches have opposite strengths and weaknesses. The iterated method produces efficient parameter estimates if the model is correctly specified but is prone to misspecification. On the contrary, the direct method is robust to misspecification but it delivers nosier estimates of the multiplier if the true DGP is approximated sufficiently well by a dynamic model like a VAR.

In this work we propose a methodology that attempts to strike a balance between LP and VARs. We work within the LP framework under the additional assumption that the coefficients in (1) are a smooth function of the forecast horizon. Such an assumption can be justified from the fact that most dynamic models, whether statistical models (e.g., VAR model) or models grounded in economic theory (e.g., DGSE models), produce multipliers that are indeed smooth function of the forecast horizon.

In order to carry out inference on the model in (1) under our smoothness assumption we resort to semi-parametric estimation methods, and specifically penalized B-splines. Penalized B-splines smoothing was popularized in statistics by Eilers and Marx (1996), and it consists in (i) approximating the unknown function of interest using a linear combination of B-splines basis functions and (ii) estimating the parameters of this linear combination via shrinkage. As we illustrate in this section, penalized B-splines can be particularly attractive in our context.

2.1.1 Semi-parametric Modeling using B-splines

We begin by approximating the coefficient $\beta_{(h)}$ using a linear B-splines basis function expansion in the forecast horizon h, that is

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

where $B_k : \mathbb{R} \to \mathbb{R}$ for $k = 1, ..., K$ is a set of B-spline basis functions and b_k for $k = 1, ..., K$ is a set of scalar parameters. We proceed analogously with the other coefficients $\alpha_{(h)}$ and $\gamma_{i\,(h)}$, so that the LP of (1) can be approximated as

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

The top panel of Figure 1 shows the set of B-splines basis functions used throughout this work. B-splines are a basis of humped 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. Moreover, the B-splines 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, the bottom panel of Figure 1 shows the B-splines basis of knot 6 together with the inner knots used to construct this function. In this work we will use a cubic B-splines basis with equidistant knots ranging from -2 to $H-1$ with unitary increments.[6] Eilers and Marx (1996) contains code for computing the B-spline basis in MATLAB and R.

A few comments are in order. First, B-splines have good approximation properties, are simple to compute and lead to a well conditioned regression model (unlike, for instance, classic splines). Second, as it is customary in the smoothing literature, we use a relatively large number of terms in the linear basis expansion in (2) to ensure that any sufficiently smooth function can be approximated with adequate accuracy. For notation simplicity we are using the same set of basis functions and number of knots for all variables but it is straightforward to generalize this if needed. Last, while here we are interested in smoothing all the coefficients of (1), one can choose to smooth only a subset of the coefficients by appropriately adjusting (3).

An appealing feature of the projection in (3) is that it retains linearity with respect to the parameters, so that standard linear regression methods can be applied for estimation (see e.g. White, 2006). To this extent, it is convenient to introduce the matrix representation of the approximated model (3). Denote by θ the vector of B-splines parameters $(a_1, ..., a_k, b_1, ...b_k, c_{11}, ..., c_{1k}, ..., c_{p1}, c_{pk})'$. Let $\mathcal{Y}$ be a $\left(T(H+1) - \frac{H(H+1)}{2}\right) \times 1$ vector with the $(1+(t-1)(H+1)+h)$ entry defined as y_{t+h} for each $t = 1, ..., T$ and $h = 0, ..., H$. Let $\mathcal{X}_\beta$ be a $\left(T(H+1) - \frac{H(H+1)}{2}\right) \times K$ matrix with the $(1+(t-1)(H+1)+h, k)$ element defined as $B_k(h)x_t$ for each $t = 1, ..., T$, $h = 0, ..., H$ and $k = 1, ..., K$. Let $\mathcal{X}_\alpha$, $\mathcal{X}_{\gamma_i}$ for $i = 1, ..., p$ be defined analogously to $\mathcal{X}_\beta$. Last, let $\mathcal{X}$ be the $\left(T(H+1) - \frac{H(H+1)}{2}\right) \times K(2+P)$ matrix obtained by stacking horizontally the matrices $\mathcal{X}_\alpha$, $\mathcal{X}_\beta$ and $\mathcal{X}_{\gamma_i}$ $i = 1, ...p$, that is

$$\mathcal{X} = \begin{bmatrix} \mathcal{X}_\alpha & \mathcal{X}_\beta & \mathcal{X}_{\gamma_1} & ... & \mathcal{X}_{\gamma_p} \end{bmatrix}.$$

Then, (3) can be compactly be represented as a linear model

$$\mathcal{Y} = \mathcal{X}\theta + \mathcal{U}, \tag{4}$$

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

where $\mathcal{U}$ denotes the vector prediction error term of the regression.

2.1.2 Penalized Estimation

The model of Equation (4) is typically expensively parameterized, and the least squares estimator may suffer from excessive variability. To address this issue, it is common practice in the smoothing literature to resort to shrinkage estimation. In particular in this work we resort to generalized ridge estimation, that is we estimate θ by minimizing the penalized residual sum of squares given by

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

where λ is a positive shrinkage parameter determining the amount of shrinkage and $\mathbf{P}$ is a symmetric positive semi-definite penalty matrix. The shrinkage coefficient λ determines the bias/variance tradeoff of the estimator: When λ is zero the estimator coincides with the least square estimator (zero bias but potentially large variance) whereas when λ is large the estimator is biased but has smaller variance than the least squares estimator. The behaviour of the generalized ridge estimator also depends on the choice of the penalty matrix $\mathbf{P}$. For example, if $\mathbf{P}$ is the identity matrix large values of λ shrink the B-spline parameter estimates towards zero.

A few comments on the practical implementation of the estimator are in order. First, to implement the estimator one has to chose the shrinkage parameter λ in (5). There is a large number of data-driven criteria for the choice of the shrinking parameter and here we resort to k-fold cross validation (see also Racine, 1997). Second, the dimensionality of the parameter θ can be large in empirical applications. However, an appealing feature of the penalized B-splines framework used in this work is that the matrices $\mathcal{X}'\mathcal{X}$ and $\mathbf{P}$ are band matrices, making the computation of the estimator numerically straightforward using appropriate algorithms designed for these type of matrices.

One of the highlights of the B-splines framework is that generalized ridge regression can be used with a family of penalty matrices $\mathbf{P}$ that allows to shrink the estimated dynamic multiplier towards a polynomial of a given order. To illustrate this feature, we consider for simplicity a constrained version of the model in (1) in which no intercept and no control variables are present and the only parameter that has to be estimated is the dynamic multiplier $\beta_{(h)}$. As it is well known in the B-splines smoothing literature (Eilers and Marx, 1996), the derivatives of order r of the approximated multiplier $\sum_{k=1}^{K} b_k B_k(h)$ with respect to the horizon h can be expressed as a linear combination of the finite differences of order r of adjacent B-splines coefficients b_k. This makes it possible to control for the degree of smoothness of the multiplier by appropriately constraining the finite differences of adjacent B-splines coefficients. In particular this motivates to estimate the model using the penalty function

$$\lambda \sum_{k=r+1}^{K} \left(\Delta^r b_k\right)^2 , \tag{6}$$

where Δ^r denotes the difference operator of order r. A large value of λ in (6) leads to shrinking the finite order differences of order r to zero, and thus the derivative of order r to zero. This in turns shrinks the estimated dynamic multiplier to a polynomial of order $r-1$. The attractive feature of the penalty in (6) is that it can be used within the generalized ridge regression framework. Let $\mathbf{D}_r$ denote the matrix representation of the r-th difference operator Δ^r.[7]Then the penalty can be expressed as

$$\lambda \sum_{k=r+1}^{K} \left(\Delta^r b_k\right)^2 = \lambda(\mathbf{D}_r b)'(\mathbf{D}_r b) = \lambda b'\mathbf{D}_r'\mathbf{D}_r b \ ,$$

so that the penalty in (6) can be imposed in the estimator defined in (5) by simply setting $\mathbf{P} = \mathbf{D}_r'\mathbf{D}_r$. It is straightforward to generalize this estimation strategy for all coefficients the model in (3). In order to shrink the B-splines approximation of

[7]For instance, with $r = 2$, we have

$$\mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & ... & & & \\ 0 & 1 & -2 & 1 & 0 & ... & & \\ ... & 0 & 1 & -2 & 1 & 0 & ... & \\ & ... & 0 & 1 & -2 & 1 & 0 & \\ & & ... & 0 & 1 & -2 & 1 & \end{bmatrix} .$$

all the $\alpha_{(h)}, \beta_{(h)}, ..., \gamma_{p\,(h)}$ coefficients to, respectively, polynomials of order $r_1 - 1, r_2 - 1, .., r_{2+p} - 1$ the penalty matrix $\mathbf{P}$ may be set to a block diagonal matrix with matrices $\mathbf{D}'_{r_1}\mathbf{D}_{r_1}, \mathbf{D}'_{r_2}\mathbf{D}_{r_2}, ..., \mathbf{D}'_{r_{2+p}}\mathbf{D}_{r_{2+p}}$ on the diagonal.

It is important to emphasize that one can easily construct penalty matrices that allow for further shape restrictions. Here we just mention one extension of interest in our context. Consider again the restricted version of model (1) comprising of the x_t term only. It may be reasonable in some applications to assume that the dynamic multiplier $\beta_{(h)}$ vanishes at a large enough horizon, for instance because there are economic reasons to believe that the effect of the shock dies out at long horizons. To this extent one may want to shrink the multiplier towards a $r - 1$ degree polynomial constrained to be zero at, say, horizon H. It is straightforward to check that one can set such a shrinking target by setting $\mathbf{P}$ to

$$\mathbf{P} = \begin{bmatrix} \mathbf{D}_r \\ e'_K \end{bmatrix}' \begin{bmatrix} \mathbf{D}_r \\ e'_K \end{bmatrix} ,$$

where e_K is a $K \times 1$ vector with a one in position K and zero otherwise.

Last, we discuss how to construct confidence bands for the dynamic multipliers of the SLP. Confidence bands for flexible models like the one considered in this work can be challenging to construct (Härdle, 1990). Moreover, as Chen (2007) points out, the asymptotic distribution theory for semi-parametric estimators needed to construct such intervals has only been partially developed in the literature so far (see also Andrews, 1991; Huang, 2003). Here, we construct confidence intervals using the following approach. The dynamic multipliers in (3) are a linear combination of the elements of θ, which makes the computation of their standard errors straightforward given an estimate of the variance of $\hat{\theta}$. As in Jordá (2005), we use HAC methods to estimate the variance covariance matrix of the $\hat{\theta}$ estimator. One of the challenges in constructing the confidence intervals lies in the fact that the ridge estimator is biased which in turns delivers a biased estimator of the multiplier. To this extent, the confidence interval of the estimated dynamic multipliers are based on an undersmoothed estimate of θ to reduce the extent of the bias.

2.2 Estimating Structural Impulse Responses

One of the main goals of macroeconomic analysis is carrying out inference on the dynamic effects of structural shocks. In this section we consider y_t as an endogenous variable in a macroeconomic system of interest and we are concerned with the estimation of its response to a structural shock ε_t. We follow the definition of Ramey (2016) and define a structural shock ε_t as a variable (i) which is exogenous with respect to the other current lagged endogenous variables in the system, (ii) is uncorrelated with other exogenous shocks and (iii) represents either unanticipated movements in exogenous variables or news about future movements in exogenous variables (see also Blanchard and Watson (1986), Bernanke (1986) and Stock and Watson (2016)). The structural IR of y_t to a structural shock ε_t is then defined as[8]

$$\mathsf{IR}(h,\delta) = \mathrm{E}(y_{t+h}|\varepsilon_t = \delta) - \mathrm{E}(y_{t+h}|\varepsilon_t = 0), \quad h = 0, ..., H \ .$$

In this section, we show how LP may be used to recover structural IRs when the structural shock (i) is observed, (ii) is not observed and it can be recovered through control variables, or (iii) is not observed and it can be recovered through instrumental variables.

2.2.1 Observed Structural Shock

If the structural shock ε_t is observed then the $\mathsf{IR}(h,\delta)$ can be obtained from the LP of y_t on ε_t (without having to include any additional regressors), that is

$$y_{t+h} = \alpha_{(h)} + \beta_{(h)}\varepsilon_t + u_{(h)\,t+h} \ .$$

The coefficient $\beta_{(h)}$ captures the causal effect of the structural shock ε_t and the IR is given by $\mathsf{IR}(h,\delta) = \beta_{(h)}\delta$, which can readily be estimated as $\widehat{\mathsf{IR}}(h,\delta) = \hat{\beta}_{(h)}\delta$. This approach has recently been used by Auerbach and Gorodnichenko (2012) and Tenreyro and Thwaites (2016) using the narratively identified monetary shock of Romer and Romer (2004). Notice

[8]Note that our definition is closely related to the "Mean of generalized impulse response function conditional on shock" defined in Koop, Pesaran, and Potter (1996).

that additional regressors may be included to "mop up" the residual variance of the regression and obtain more precise estimates (Ramey, 2016).

2.2.2 Identification through Controls

Unfortunately, structural shocks are seldom observable. However, it is sometimes possible to recover structural shocks as the innovation of an endogenous variable in the system conditional on an appropriate set of controls. Specifically, in some cases, the structural shock ε_t is given by the residual of the endogenous variable x_t conditional on a set of control variables, namely

$$\varepsilon_t = x_t - \mathrm{E}(x_t | w_{1\,t}, ..., w_{p\,t}) \,, \tag{7}$$

and that this conditional expectation is linear in $w_{i\,t}$ for $i = 1, ...p$. Then, the $\beta_{(h)}$ coefficient of the LP of y_{t+h} on x_t and $w_{i\,t}$ for $i = 1, ..., p$ in equation (1) captures the effect of the structural shock ε_t (see also Angrist, Jordá, and Kuersteiner, 2016; Jordá and Taylor, 2016).

The recursive identification scheme put forward in Sims (1980) may be seen as special case of such an approach. Sims (1980) proposes timing assumptions among the exogenous shocks in a VAR to disentangle the causal chain of events and identify structural shocks of interest. In the LP setting, the timing restrictions correspond to a specific choice of control variables. We illustrate this with an example below, and we provide a formal proof in the appendix.

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 $y_{t+h} = \mathsf{gdp}_{t+h}$, $x_t = \mathsf{ffr}_t$ and $\boldsymbol{w}_t' = (\mathsf{gdp}_t, \pi_t, \mathsf{gdp}_{t-1}, \pi_{t-1}, \mathsf{ffr}_{t-1})'$. Intuitively, we achieve identification by controlling for the contemporaneous values of variables ordered before the shock of interest (in this case, output and inflation).

Importantly, note that the LP approach allows to use recursive identification beyond linear VAR to nonlinear DGPs provided that the set of variables is appropriately augmented. For instance, the set may possibly have to include nonlinear transformation of the variables in the system.

2.2.3 Identification through Instruments

Even when condition (7) fails, it may still be possible to recover the structural shock of interest if an appropriate instrumental variable is available. For instance, when the set of control variables does not allow to identify the structural shock of interest, we may have that

$$\varepsilon_t + e_t = x_t - \mathrm{E}(x_t | w_{1\,t}, ..., w_{p\,t}) \,,$$

where e_t is a serially uncorrelated measurement error.

In these cases an instrument can allow to recover the effect of interest. Specifically, we define an instrument z_t to be a time-series satisfying

$$\mathrm{corr}(\varepsilon_t, z_t) \neq 0 \text{ and } \mathrm{corr}(e_t, z_t) = 0, \tag{8}$$

in other words z_t is assumed to be *relevant* and *exogenous*. Then, we can recover the IR of y_t to a structural shock ε_t from (S)LP (1) using z_t as an instrument.

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). The example of Section 2.2.1, assumed the Romer and Romer (2004) shock series to be a perfect measurement of the true monetary shock. A more reasonable assumption however may be to posit that the Romer and Romer (2004) shocks can only proxy for the true monetary shocks, 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 (8), 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 $y_{t+h} = \mathsf{gdp}_{t+h}$, $x_t = \hat{\mathsf{ffr}}_t$ where

$\hat{\mathsf{ffr}}_t$ is the fitted value of the first stage regression. Then, the $\beta_{(h)}$ coefficient captures the effect of a monetary shock on output. As previously discussed in Section 2.2.1 additional regressors may be included to "mop up" the residual variance of the regression.

Our proposed instrumental approach parallels the "external" instruments approach pioneered by Stock and Watson (2008) and Mertens and Ravn (2013) in the context of VAR identification and referred to as "proxy SVAR" by Ramey (2016). Our proposed approach allows to take advantage of all the appealing features of (S)LP within the instrumental variable framework.

2.3 Estimating State-dependent Structural Impulse Responses

In this section, we discuss the identification and estimation of non-linear IRs with (S)LP, and more specifically state-dependent IRs, which have recently been the subject of important interest in macroeconomics.[9]

A significant number of contributions focus on proposing methodologies to estimate state-dependent IRs. These contributions typically rely on regime-switching models, in which a state variable (observed or not) controls the dynamics of the economy (e.g., Hamilton, 1989, Hubrich and Terasvirta, 2013). These model can be expensively parametrized (allowing for only a small number of states) and computationally intensive to estimate.

(S)LP offer an alternative approach to estimate state-dependent IRs. To capture state-dependence, we allow $\beta_{(h)}$ to depend linearly on the value of an observed state variable s_t, that is

$$\beta_{(h)} = \beta_{0\,(h)} + \beta_{1\,(h)} s_t \ .$$

Note how the state variable affects the IR in a continuous fashion. This is in contrast to regime-switching models where there is a discrete (and small) number of states with a separate IR for each state.

In what follows the $\beta_{1\,(h)}$ coefficient capturing the amplification/contraction effect due to the state variable s_t is called state multiplier. Note that in this case the model in (1)

[9]See e.g, Auerbach and Gorodnichenko (2012), Ramey and Zubairy (2014b), Jordá and Taylor (2016), Tenreyro and Thwaites (2016) and Santoro *et al.* (2014)

becomes

$$
\begin{aligned}
y_{t+h} &= \alpha_{(h)} + \left(\beta_{0\,(h)} + \beta_{1\,(h)} s_t\right) x_t + \sum_{i=1}^{p} \gamma_{i\,(h)} w_{i\,t} + u_{(h)\,t+h} \\
&= \alpha_{(h)} + \beta_{0\,(h)} x_t + \beta_{1\,(h)} x_t s_t + \sum_{i=1}^{p} \gamma_{i\,(h)} w_{i\,t} + u_{(h)\,t+h}
\end{aligned}
$$

which can still be approximated using standard (S)LP. In particular, the model can be estimated using the SLP methodology outlined in the previous sections.

Importantly, identification of the structural IRs proceeds similarly to the linear case. We provide a formal proof in the appendix where we show how one can use LP to consistently estimate (non-linear) structural IRs when the DGP is a vector moving-average whose coefficients satisfy $\beta_{(h)} = \beta_{0,(h)} + \beta_{1,(h)} s_t$. In this case, the set of controls $w_{i\,t}$ has to include those of the linear case plus their interaction with the state variable.

To give a concrete application, consider our previous monetary example identified with a recursive identifying scheme. We are interested in knowing whether monetary policy has a larger effect on output during recessions. Taking our state variable indicator s_t to be the level of GDP growth at the time of the shock, one can recover the state-dependent IR of GDP from the LP by setting $y_{t+h} = \mathsf{gdp}_{t+h}$, $x_t = \mathsf{ffr}_t$ and $\boldsymbol{w}_t' = (\mathsf{gdp}_t, \pi_t, \mathsf{gdp}_t s_t, \pi_t s_t, \{\boldsymbol{\chi}_{t-k}\}_{k=1}^{K}, \{\boldsymbol{\chi}_{t-k} s_{t-k}\}_{k=1}^{K})'$ where $\boldsymbol{\chi}_t = (\mathsf{gdp}_t, \pi_t, \mathsf{ffr}_t)'$.

3 Simulation Study

In this section we carry out a simulation study to illustrate our proposed methodology as well as to benchmark the performance of IR estimation based on SLP against (regular) LP and SVARs.

We consider a system comprising GDP growth gdp_t, inflation π_t and fed funds rate

ffr_t. The trivariate system $(\mathsf{gdp}_t, \pi_t, \mathsf{ffr}_t)'$ is generated as

$$\begin{aligned}
\mathsf{gdp}_t &= \sum_{h=0}^{H} \beta_{11\,(h)} \varepsilon^{\mathsf{gdp}}_{t-H+h} + \sum_{h=1}^{H} \beta_{12\,(h)} \varepsilon^{\pi}_{t-H+h} + \sum_{h=1}^{H} \beta_{13\,(h)} \varepsilon^{\mathsf{ffr}}_{t-H+h} \\
\pi_t &= \sum_{h=0}^{H} \beta_{21\,(h)} \varepsilon^{\mathsf{gdp}}_{t-H+h} + \sum_{h=0}^{H} \beta_{22\,(h)} \varepsilon^{\pi}_{t-H+h} + \sum_{h=1}^{H} \beta_{23\,(h)} \varepsilon^{\mathsf{ffr}}_{t-H+h} \\
\mathsf{ffr}_t &= \sum_{h=0}^{H} \beta_{31\,(h)} \varepsilon^{\mathsf{gdp}}_{t-H+h} + \sum_{h=0}^{H} \beta_{32\,(h)} \varepsilon^{\pi}_{t-H+h} + \sum_{h=0}^{H} \beta_{33\,(h)} \varepsilon^{\mathsf{ffr}}_{t-H+h}
\end{aligned} \tag{9}$$

where $\varepsilon^{\mathsf{gdp}}_t$, ε^{π}_t and $\varepsilon^{\mathsf{ffr}}_t$ are i.i.d. structural normal shocks with mean zero and variances equal to, respectively, σ^2_{gdp}, σ^2_{pi} and σ^2_{ffr}. The horizon parameter H is set to 20. Notice that standard timing restrictions are imposed on the contemporaneous impact of the shocks in the system. In particular, GDP growth shocks have a contemporaneous impact on all series in the system, inflation shocks have a contemporaneous impact on inflation and the fed funds rate, and the fed funds rate/monetary shocks have a contemporaneous impact on the fed funds rate only. These restrictions allow to identify the full set of structural impulse response functions in the system using standard methods (see Section 2.2). Throughout this section the focus is on the estimation of the IRs of gdp_t and π_t to a monetary shock $\varepsilon^{\mathsf{ffr}}_t$ from the model of Equation (9) up to horizon $H = 20$ using SLP, LP and VARs.

INSERT FIGURE 2 ABOUT HERE

We consider two different parameter settings for the model of Equation 9 labelled as *rough* and *smooth*. In order to entail realistic data dynamics, the DGPs considered in this study are based on estimates obtained from a quarterly macro dataset comprising of GDP growth, PCE inflation and the fed funds rate from from 1959-Q1 to 2007-Q4.[10] In the rough setting, the parameters of model (9) are obtained from the estimates of the nine structural IRs obtained by LP from our reference macro dataset. Specifically, we identify the IRs of the structural shocks in (9) through controls (see Section 2.2.2) by including in

[10] We exclude the latest recession as the fed funds rate was constrained at zero and no longer captured variations in the stance of monetary policy.

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 $\boldsymbol{w}_t' = (\mathsf{gdp}_t, \mathsf{gdp}_{t-1}, \pi_{t-1}, \mathsf{ffr}_{t-1}, \ldots, \mathsf{gdp}_{t-4}, \pi_{t-4}, \mathsf{ffr}_{t-4})'$. In the smooth setting, we use the same parameters as in the rough one except that we use smoothed versions of the dynamic multipliers of GDP growth and PCE inflation to a fed funds rate shock. These smoothed multipliers are obtained by regressing the LP estimates on a sine/cosine basis

$$\hat{\beta}_{ij\,h} = c_1 \sin\left(\frac{2\pi}{H}h\right) + c_2 \cos\left(\frac{2\pi}{H}h\right) + c_3 \sin\left(\frac{2\pi}{H}2h\right) + c_4 \cos\left(\frac{2\pi}{H}2h\right) + u_h \ ,$$

for $i = 1, 2$ and $j = 3$ and then using the fitted values of the regression as the new set of dynamic multipliers.[11] Figure 2 plots the IRs of gdp_t and π_t used in respectively the rough and the smooth settings. In the rough case IRs can exhibit abrupt changes across horizons whereas in the smooth case IRs change smoothly across horizons.

We estimate the IRs of gdp_t and π_t to a monetary shock using SLP, LP, and a SVAR using timing restrictions consistent with our DGP. For all three estimators we set the number of lags to 4.

A number of details on the implementation of the SLP estimator used in this study are in order. First, we only impose smoothness on the coefficients associated with the IR and we do not smooth the coefficients of the control variables. It is important to emphasize that this allows us to compare more easily the LP and SLP estimators but makes the exercise more disadvantageous for our SLP methodology as further efficiency gains could be attained by imposing smoothness on the remaining coefficients. As far as the choice of the penalty matrix is concerned, here we opt for a naïve approach and shrink towards a polynomial roughly consistent with the IR estimated by the standard LP. We shrink towards a line for GDP growth and towards a quadratic polynomial for inflation, that is, we penalize the SLP using a difference penalty of order $r = 2$ for GDP and $r = 3$ for inflation. Last, the shrinkage parameter λ is chosen by 5-fold cross-validation. For

[11] Note that we use a different smoothing method than P-splines in order to not bias our results in favor of SLP.

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 of the target IR of interest. The Oracle shrinkages level is determined by simulation.

INSERT FIGURE 3 ABOUT HERE

3.1 Illustration

We use one simulation from the smooth DGP to illustrate some aspects of our SLP methodology. In this example, the sample size of our simulated system is 200.

Figure 3 shows the IR estimates based on SLP (based on cross-validation), LP and SVAR for the two IRs of interest. Note that despite the population IRs being smooth, the LP delivers IR estimates that are quite rough, a well known feature of LP (e.g. Ramey, 2016). We can see that SLP essentially smooths the LP, and in this particular replication, delivers a more precise estimate of the IRs. Last, the VAR IRs deviates substantially from the true IR at medium/long horizon in both cases.

INSERT FIGURE 4 ABOUT HERE

Figure 4 shows how the SLP IR estimates change as a function of the shrinkage parameter 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 other words, SLP nests two important IR estimation methods: When the penalty parameter is set to zero, SLP reduces to a (nonparametric) standard LP, and when the penalty parameter is large, SLP reduces to a (parametric) Almon distributed lag model, in which the IR is given by a polynomial function.

INSERT FIGURE 5 ABOUT HERE

The top two panels of Figure 5 show how the choice of the penalty matrix determines the limiting behaviour of the SLP IR estimator. We consider different target polynomials with r ranging from 0 to 3, which correspond to the following target polynomials: A zero function, a constant, a line and a parabola. To obtain the specific polynomial implied by the data, we estimate SLPs using a large amount of shrinkage ($\lambda = 10^6$). The figure shows that when λ is large the SLP estimator collapses to the best fitting polynomial of order $r - 1$. As we increase the order of the polynomial, the target polynomial adjusts and allows for richer shapes of the target IR. The bottom two panels of Figure 5 consider a refined penalty matrix in which we impose that the target polynomial equals zero at horizon 20 as explained in Section 2.1.2.

3.2 Simulation Results

We replicate our simulation exercise for the rough DGP and smooth DGP using a sample size equal to 50, 100, 200, 300 and 400. The simulation is replicated 1000 times for each setting and sample size. The performance of each IR estimator is measured by its integrated MSE defined as

$$\mathsf{MSE} = \mathrm{E}\left[\sum_{h=0}^{H} (\widehat{\mathsf{IR}}(h,\delta) - \mathsf{IR}(h,\delta))^2\right],$$

which is approximated using the Monte Carlo average across replications.

INSERT TABLE 1 AND FIGURE 6 ABOUT HERE

Table 1 reports the average MSE of VAR, LP, SLP and Oracle SLP. Results show that SLP out-performs LP in both the rough and smooth setting. Although perhaps surprising, the superior performance of SLP in the rough setting comes from the benefits of shrinkage regression.[12] The SLP and LP also perform better than the VAR approach provided that the sample size is larger than 50 observations.

In order to inspect the simulation results more easily, in Figure 6 we plot the MSE of each estimator as a function of sample size for the case of the smooth DGP. A number of comments are in order.

[12] In other words, in these simulations, the estimation variance of LP is such an acute problem in short samples that shrinkage delivers a lower MSE even if the underlying DGP is not smooth.

First, we can see that SLP performs better than LP for all sample size, but also that the improvement is largest for small samples. This result is to be expected, since SLP precisely helps IR estimation by reducing the estimation variance through shrinkage. In larger samples, estimation variability is less of a concern, and the advantage of SLP over LP diminishes. However, for samples with $T = 50$ or 100, the large gains offered by SLP are particularly promising, since such relatively sample sizes are often encountered in macro applications.[13]

INSERT FIGURE 7 ABOUT HERE

Second, comparing SLP with the VAR, we can see that while SLP dominates in large samples, the VAR delivers better performances in shorter samples (when $T = 50$). To better understand this result, we decompose the total MSE of each IR estimator into the contribution of the different horizons and in Figure 7 we plot $\mathrm{E}\left[(\widehat{\mathsf{IR}}(h, \delta) - \mathsf{IR}(h, \delta))^2\right]$. The top panels plot the decomposition for a sample size of $T = 50$ and the bottom panels for $T = 400$. Detailed inspection of the results reveals that the VAR IRs have a bias which is particularly severe for both IRs from horizon 12 to 16 (especially in the case of inflation). This is apparent in the figures where the VAR MSE exhibits noticables bumps for both IRs. When the sample size is small, the parametric nature of VAR produces better results relatiave to (S)LP despite its bias. On the other hand, when the sample is large enough, the bias of the VAR leads to a significant performance deterioration with respect to (S)LP.

Finally, comparing the performances of the Oracle SLP with the cross-validated SLP we see that there are no large differences between using the optimal λ and selecting a λ from cross-validation, hinting that, for the class of DGPs considered in this study, cross-validation performs satisfactorily.

[13]This is often the case with non-US data. For instance, empirical work with eurozone data is limited to about 15 years of data, that is 60 quarters.

4 Empirical Application

In this section we use our proposed methodology to study the effects of monetary shocks, which have been the subject of extensive research (see Ramey (2016) for a review). Here we apply our SLP approach using the two main identification schemes used in the literature: (i) identification based on timing restrictions and (ii) identification using instrument variables.

4.1 The Linear Effects of Monetary Shocks

We consider a system comprising of GDP growth, PCE inflation and federal funds rate. The dataset spans from 1959-Q1 to 2007-Q4. We exclude the latest recession where the fed funds rate was constrained at zero and no longer captured variations in the stance of monetary policy.

We first identify the IR of monetary shocks using timing restrictions on the base of the argument that monetary policy affects macro variables with a lag. More precisely, we assume that that we can identify the IR of GDP growth (inflation) to a monetary shock from a (S)LP of GDP growth (inflation) on the fed funds rate using as controls the contemporaneous value of GDP growth and inflation as well as 4 lags of GDP growth, inflation and the fed funds rate.

Figure 8 plots the IRs of GDP growth and inflation to a one standard-deviation monetary shock. The top panel plots the impulse responses obtained from LPs, while the bottom panel plots the IRs obtained from SLP. Following a contractionary shock, GDP growth and inflation decline, as previously found in numerous studies. Note however, that unlike previous studies, the recursive identification assumption is used directly in the local projections and no VAR is estimated.

An important advantages of SLP over LP are apparent in Figure 8. Namely, the IRs obtained by regular LP can be erratic with sometimes sharp fluctuations within a quarter as with GDP growth at $h = 3$. 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 (see e.g., Ramey (2012) for a concrete example of such a problem). In contrast, thanks to smoothing, the SLP IRs are easier to interpret. For instance, the sharp movements of GDP growth at $h = 3$ disappears under smoothing, indicating that this is not a robust feature of the data.

Next, Figure 9 plots the same IRs using an instrument variable identification scheme. As external instrument for movements in the fed funds rate, we use the Romer and Romer monetary shocks series (Romer and Romer, 2004) and extended until 2007 by Coibion *et al.* (2012). The sample spans 1966-Q1 to 2007-Q4. As controls, we include 4 lags of GDP growth, inflation and the fed funds rate. The IRs are similar to the recursive identification case, with again erratic LP estimates.

Finally, note that our SLP framework has an important additional advantage over previous approaches: it allows to implement different identifying schemes using the same methodology. In previous studies, the magnitude of the effect of monetary shocks could vary a lot across identification methods (recursive versus narrative), and in the literature on the effects of monetary shocks, the discrepancy across methods was so large that it was not clear whether the effect of monetary policy was "big or small" (Coibion, 2012).[14] One of the issue was that the recursive identification strategy was implemented in VARs (e.g. Christiano *et al.*, 1999) while the narrative identification strategy was implemented using ADL methods (Romer and Romer, 2004). In contrast, when we implement the two identification schemes using the same SLP approach, we obtain similar IR estimates for both GDP and inflation.

4.2 The State-Dependent Effects of Monetary Shocks

We now explore whether the state of the cycle can affect the effectiveness of monetary policy, a topic that has received much attention in the recent years.[15]

We use the 7-quarter centered moving-average of GDP growth as the cyclical indicator s_t, following Auerbach and Gorodnichenko (2012), and we use a recursive identification

[14] See the reconciliation study of Coibion (2012).

[15] See e.g., Tenreyro and Thwaites (2016), Santoro *et al.* (2014), Barnichon and Matthes (2016).

scheme.

Figure 10 plots the response of GDP growth and inflation to a monetary shock that takes place (i) in a recession ($s_t = -1\sigma(s_t)$), (ii) in an "average" state ($s_t = 0$) and (iii) in an expansion ($s_t = +1\sigma(s_t)$). Figure 11 plots the state multipliers obtained from LP (top row) or SLP (bottom row). Recall that the state multiplier captures the extent to which the state affects the IR at each horizon, so that a positive value implies that the IR response is more negative when the economy is in a recession (the state variable is negative).

Again, while LP delivers erratic estimates with large confidence intervals (especially for inflation), SLP delivers a clear message about the state-dependent effect of monetary policy. The effect of a contractionary shock on GDP growth is substantially larger in recessions, i.e., when growth is negative: While a contractionary monetary shock lowers GDP growth by 0.5 percentage point on average, the same shock lowers GDP growth by more than one percentage point if the shock takes place during a period of slow growth. Interestingly, our results also imply that monetary policy has almost no effect on real variables during expansions. The response of inflation is also state dependent and is weaker in recessions (the price puzzle is larger), which according to a a Keynesian narrative is consistent (at least qualitatively) with the stronger response of GDP in recessions: monetary policy will have a stronger impact on real quantities if prices react less.

5 Conclusions

This paper proposes a novel IR estimation approached based on Smooth Local Projection (SLP). SLP aim at combining the flexibility of Local Projections with the efficiency of VARs. SLP consists in using Penalized-B splines to estimate local projections while imposing smoothness on the IR coefficients. While different smoothing techniques are possible, P-splines have a number of unique advantages. First, P-splines estimation is straightforward, because it takes the form of a standard ridge regression. Second, P-splines are flexible and can be used to not only impose smoothness on the impulse response function, but also to impose shapes on the estimated IR. This is particularly attractive in macroeconomics when prior knowledge on the shape of the IR is available.

We obtain two sets of results that are of particular interests for macro-economists. First, SLP can be used with common identification schemes –timing restrictions or instrumental variables– to directly estimate structural IRs. In other words, while researchers have relied on VARs to implement identification schemes based on a Cholesky ordering or an external instrument, we show that it is possible to bypass the VAR entirely and thus reduce the risk of model mis-specification. Second, for the estimation of state-dependent IRs, SLP offers a flexible semi-parametric alternative to regime-switching models, which are expensively parametrized (allowing for only a small number of states) and computationally intensive to estimate.

To illustrate the benefits of SLP with a concrete example, we use SLP to study the effects of monetary shocks with different identification schemes and to study how the effects of monetary policy varies with the state of the economy.

References

Almon, S. (1965). The distributed lag between capital appropriations and net expenditures. *Econometrica*, **33**, 178–19.

Andrews, D. W. K. (1991). Asymptotic Normality of series estimators for nonparametric and semiparametric models. *Econometrica*, **59**, 307–345.

Angrist, J., Jordá, O., and Kuersteiner, G. (2016). Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited. *Journal of Business and Economic Statistics*.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Fiscal Multipliers in Recession and Expansion. In *Fiscal Policy after the Financial Crisis*, NBER Chapters, pages 63–98. National Bureau of Economic Research, Inc.

Barnichon, R. and Matthes, C. (2016). Gaussian mixture approximations of impulse responses and the non-linear effects of monetary shocks. Cepr working paper.

Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy*, **25**(1), 49–99.

Blanchard, O. J. and Watson, M. W. (1986). Are Business Cycles All Alike? In *The American Business Cycle: Continuity and Change*, NBER Chapters, pages 123–180. National Bureau of Economic Research, Inc.

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

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. J. Heckman and E. E. Leamer, editors, *Handbook of Econometrics*, volume 6B, chapter 76, pages 1–2. North-Holland.

Christiano, L., Eichenbaum, M., and Evans, C. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics*, **1**, 65–148.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics*, **4**(2), 1–32.

Coibion, O. Y., Gorodnichenko, L. K., and Silvia, J. (2012). Innocent Bystanders? Monetary Policy and Inequality in the US. NBER Working Papers 18170, National Bureau of Economic Research, Inc.

Eilers and Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, **11**, 89–102.

Hansen, B. (2014). Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation. In J. Racine, L. Su, and A. Ullah, editors, *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press.

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

Hastie, T., R. T. and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

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

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

Jordá, O. and Taylor, A. M. (2016). The time for austerity: Estimating the average treatment effect of fiscal policy. *The Economic Journal*, **126**(590), 219–255.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, **74**(1), 119–147.

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

Mertens, K. and Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, **103**(4), 1212–47.

Monnet, E. (2014). Monetary Policy without Interest Rates: Evidence from France's Golden Age (1948 to 1973) Using a Narrative Approach. *American Economic Journal: Macroeconomics*, **6**(4), 137–69.

Racine, J. (1997). Feasible cross-validatory model selection for general stationary processes. *Journal of Applied Econometrics*, **12**(2), 169–179.

Ramey, V. (2012). Comment on "Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment";. In *NBER Macroeconomics Annual 2012, Volume 27*, NBER Chapters, pages 147–153. National Bureau of Economic Research, Inc.

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

Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. In H. Uhlig and J. Taylor, editors, *Handbook of Macroeconomics*. Elsevier.

Ramey, V. A. and Zubairy, S. (2014a). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. NBER Working Papers 20719, National Bureau of Economic Research, Inc.

Ramey, V. A. and Zubairy, S. (2014b). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. NBER Working Papers 20719, National Bureau of Economic Research, Inc.

Romer, C. D. and Romer, D. H. (1989). Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pages 121–184. National Bureau of Economic Research, Inc.

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

Romer, C. D. and Romer, D. H. (2010). The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review*, **100**(3), 763–801.

Romer, C. D. and Romer, D. H. (2014). Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Changes, 1952-1991. NBER Working Papers 20087, National Bureau of Economic Research, Inc.

Romer, C. D. and Romer, D. H. (2015). New Evidence on the Impact of Financial Crises in Advanced Countries. NBER Working Papers 21021, National Bureau of Economic Research, Inc.

Santoro, E., Petrella, I., Pfajfar, D., and Gaffeo, E. (2014). Loss aversion and the asymmetric transmission of monetary policy. *Journal of Monetary Economics*, **68**(C), 19–36.

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

Stock, J. and Watson, M. (2008). Lecture 7: Recent Developments in Structural VAR Modeling. Presented at the national bureau of economic research summer institute minicourse: What's new in econometrics: Time series, cambridge, ma, july 15.

Stock, J. H. and Watson, M. W. (2016). Factor models and structural vector autoregressions in macroeconomics. In H. Uhlig and J. Taylor, editors, *Handbook of Macroeconomics*. Elsevier.

Tenreyro, S. and Thwaites, G. (2016). Pushing on a string: Us monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics*, **8**(4), 43–74.

Walsh, C. E. (2010). *Monetary theory and policy*. The Mit Press, Massachusetts.

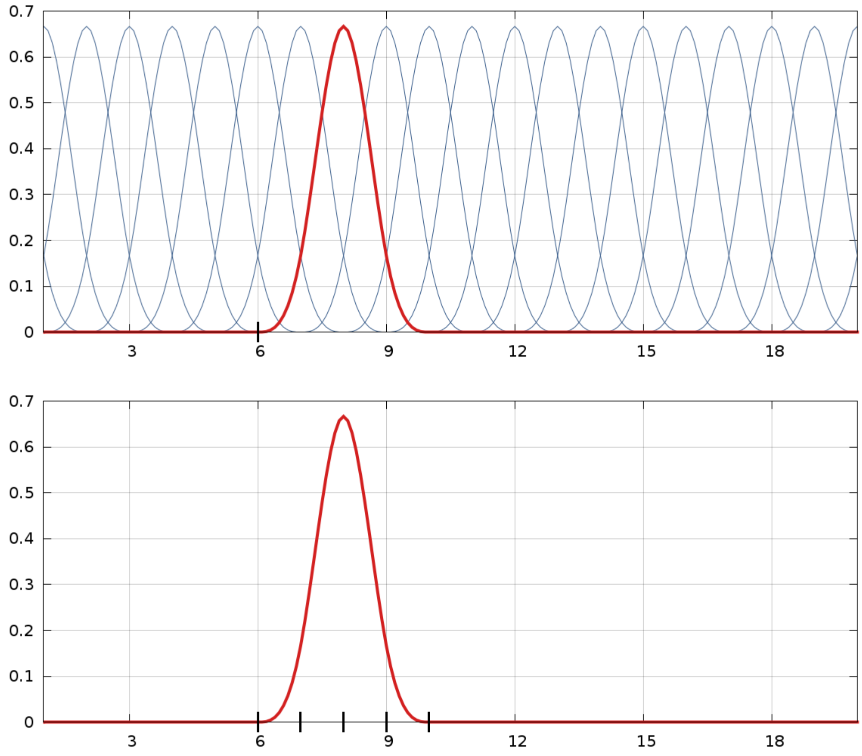
White, H. (2006). Approximate Nonlinear Forecasting Methods. In G. Elliott, C. W. J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting* , volume 1, chapter 9, pages 459–512. Elsevier.

Table 1: IR ESTIMATORS COMPARISON

	T	Rough DGP VAR	Rough DGP LP	Rough DGP SLP CV	Rough DGP SLP Oracle	Smooth DGP VAR	Smooth DGP LP	Smooth DGP SLP CV	Smooth DGP SLP Oracle
gdp_t	50	2.685	6.942	4.865	4.664	2.175	6.327	2.563	2.241
	100	2.127	3.321	2.666	2.600	1.687	3.151	1.338	1.202
	200	1.850	1.693	1.494	1.427	1.452	1.619	0.776	0.655
	300	1.787	1.200	1.088	1.047	1.369	1.130	0.577	0.465
	400	1.729	0.931	0.852	0.828	1.330	0.915	0.488	0.381
π_t	50	17.161	21.350	19.477	19.309	15.365	21.172	17.995	17.774
	100	12.576	10.069	9.388	9.302	11.632	10.146	8.711	8.595
	200	10.467	5.365	5.200	5.168	9.455	5.040	4.333	4.272
	300	9.586	3.688	3.666	3.626	8.437	3.384	2.915	2.860
	400	9.123	2.846	2.869	2.817	8.055	2.638	2.277	2.234

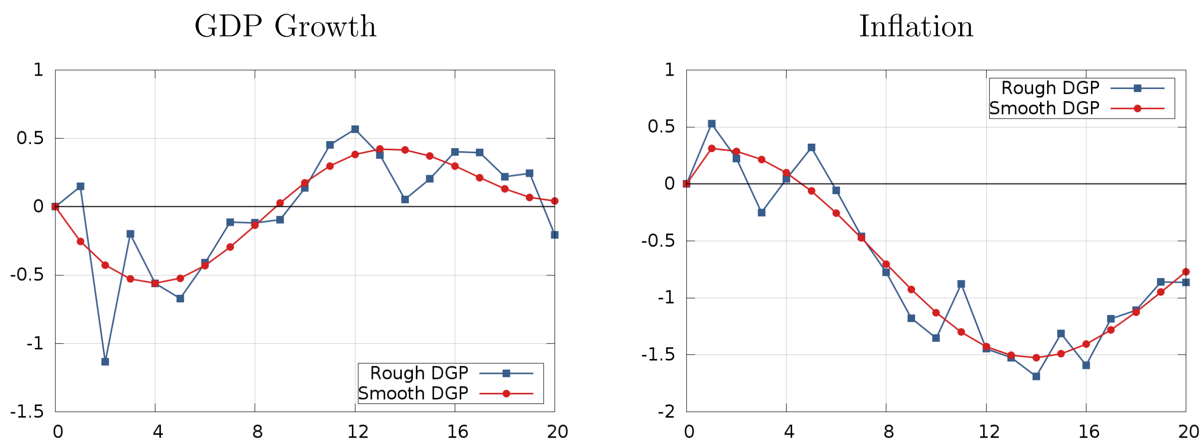
The table reports the MSE of the IR of GDP growth and inflation to a monetary policy shocks estimated via VAR, LP, SLP (based on cross-validation) and SLP (based on the Oracle shrinkage).

Figure 1: B-SPLINE BASIS



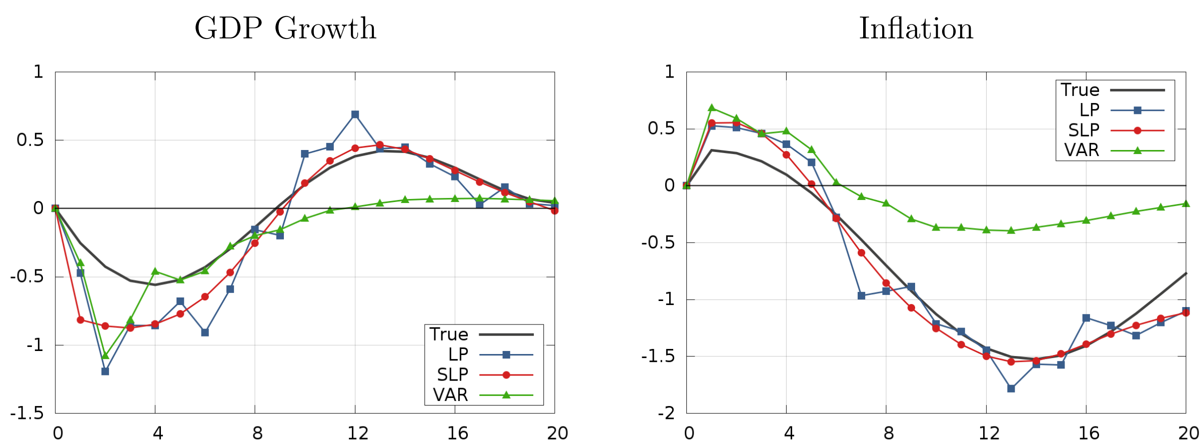
The figure shows the graph of the B-splines basis functions. The top panel displays the set of B-splines basis used in this work. The bottom panel shows in detail the B-splines basis of knot 6.

Figure 2: Simulation Study Population IRs



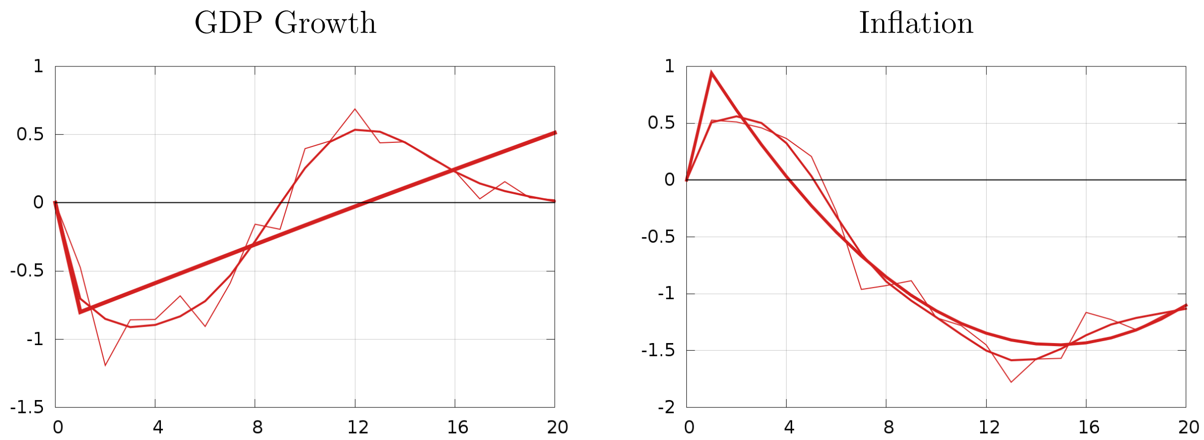
The figure displays the population IRs used in the simulation study. The left shows the IRs of GDP growth to a monetary shock in the rough DGP (squares) and smooth DGP (circles). The right shows the IRs of inflation to a monetary shock in the rough DGP (squares) and smooth DGP (circles).

Figure 3: IR estimators comparison



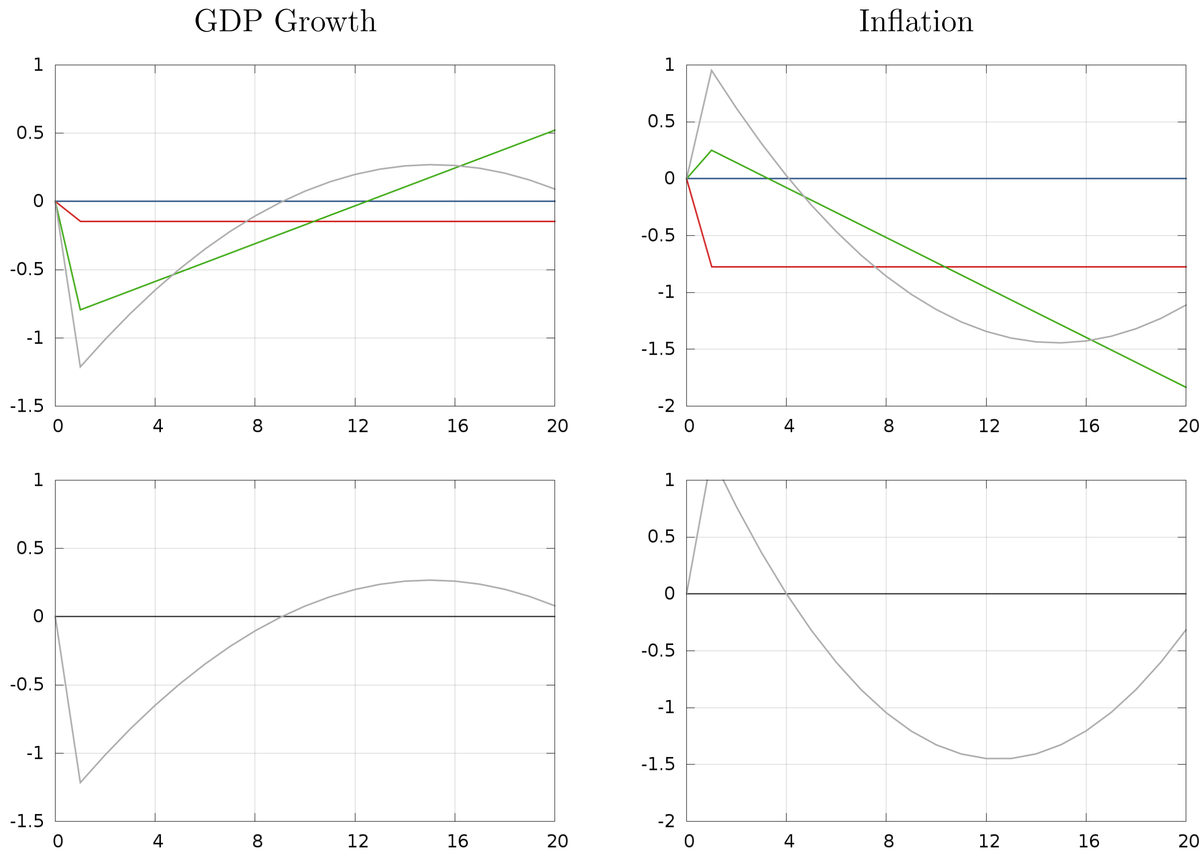
The figure displays the estimated IR in one replication of the simulation study for the smooth DGP. The panels show the estimated IR of a VAR (triangle), LP (square) and SLP (circle) as well as the population IR (solid line) for GDP growth (left panel) and inflation (right panel).

Figure 4: SLP AND THE DEGREE OF SHRINKAGE



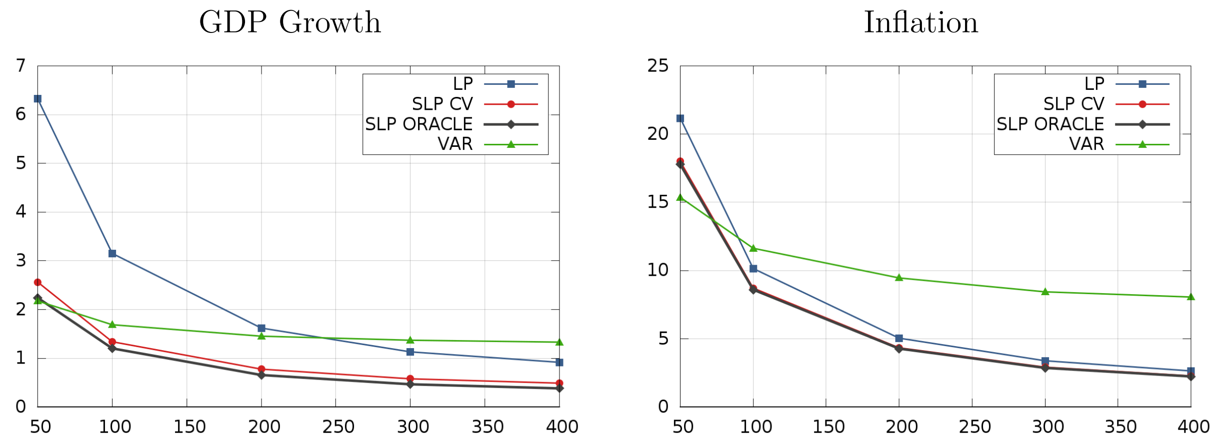
The figure displays the estimated IRs obtained from SLP using different degrees of shrinkage in one replication of the simulation study for the smooth DGP for GDP growth (left panel) and inflation (right panel). Thicker lines correspond to estimates obtained using a higher degree of penalization.

Figure 5: SLP AND TARGET POLYNOMIAL



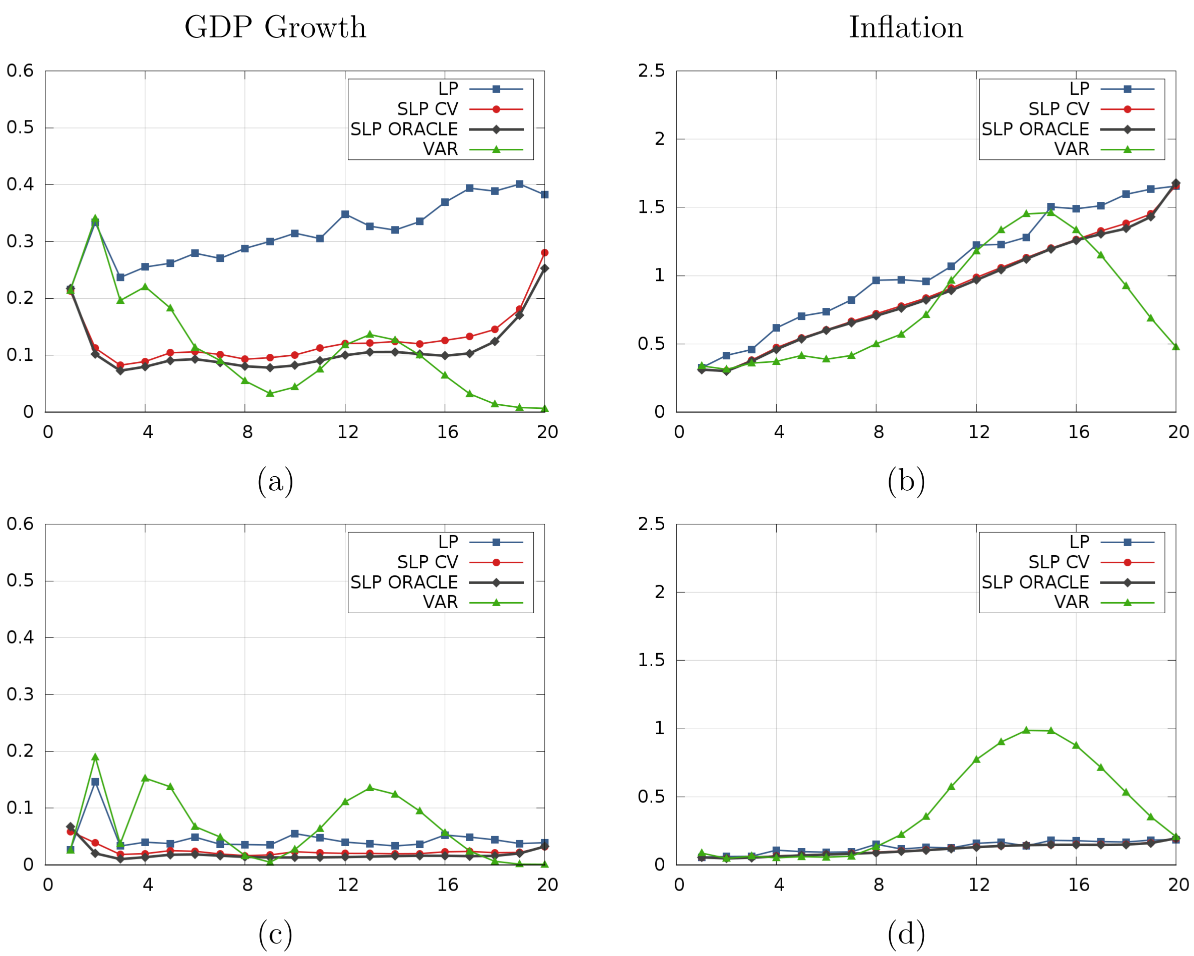
The top panels of the figure display the limiting shrinkage IR estimates for GDP growth (left panel) and inflation (right panel) obtained by setting the order of the degree penalty r to 0 (zero function) 1 (constant), 2 (line) and 3 (parabola). The bottom panels of the figure display the limiting shrinkage IR estimates for GDP growth (left panel) and inflation (right panel) obtained by setting the order of the degree penalty r to 2 and imposing shrinkage at zero at horizon 20.

Figure 6: Integrated MSE as a function of the sample size



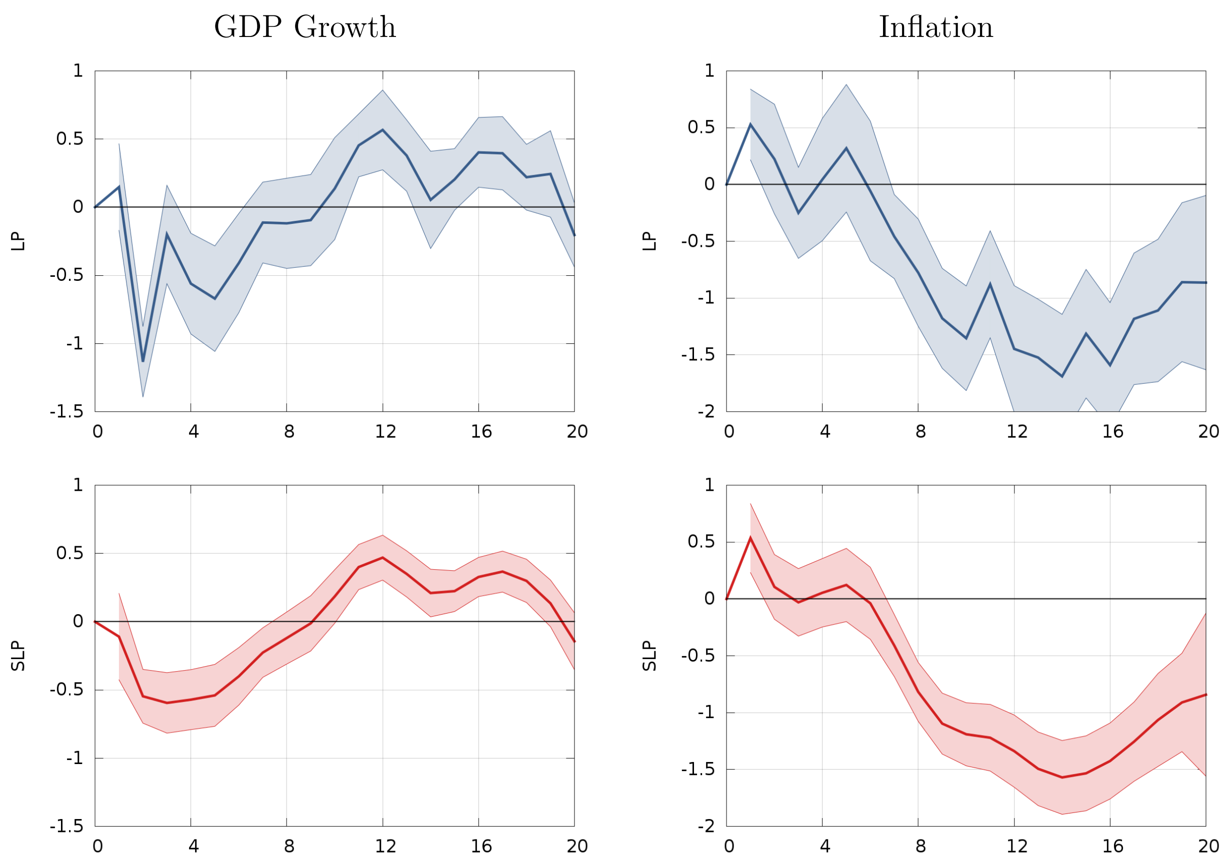
The figure displays the (integrated) MSE of the IR estimators of GDP growth (left panel) and inflation (right panel) for the smooth DGP obtained from the VAR, LP, SLP (based on cross-validation) and SLP (based on Oracle shrinkage) as a function of the sample size.

Figure 7: Point MSE as a function of the horizon



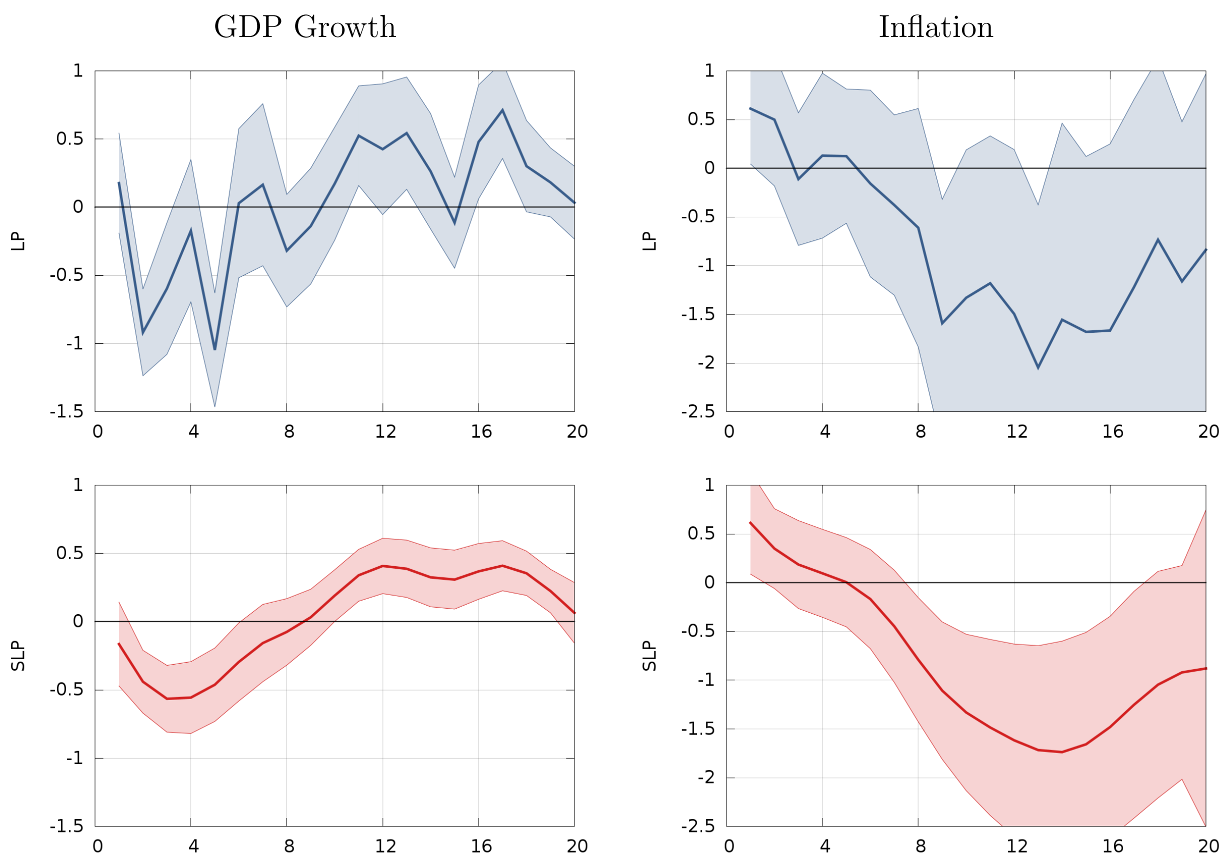
The figure displays the (pointwise) MSE of the IR estimators of GDP growth (left panels) and inflation (right panels) as a function of the horizon for the smooth DGP obtained from the VAR, LP, SLP (based on cross-validation) and SLP (based on Oracle shrinkage) as a function of the sample size. The top figures show the MSE for a sample size equal to 50 and the bottom for a sample size of 500.

Figure 8: IR TO A MONETARY SHOCK USING TIMING RESTRICTIONS



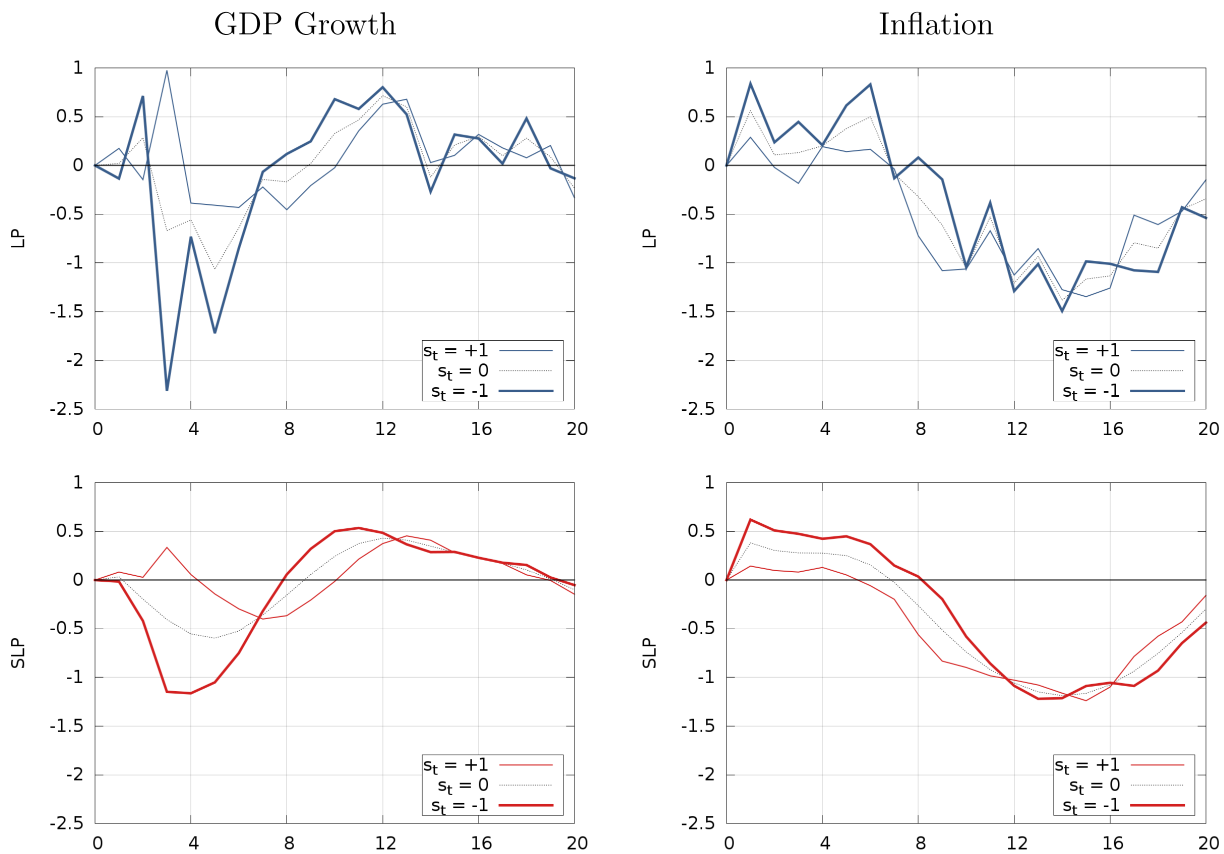
The figure displays the IR of GDP growth (left panels) and inflation (right panels) to a monetary shock identified using timing restrictions and estimated using LP (top panels) and SLP (bottom panels). The shaded area denotes the 90% confidence interval.

Figure 9: IR to a monetary shock using instrumental variables



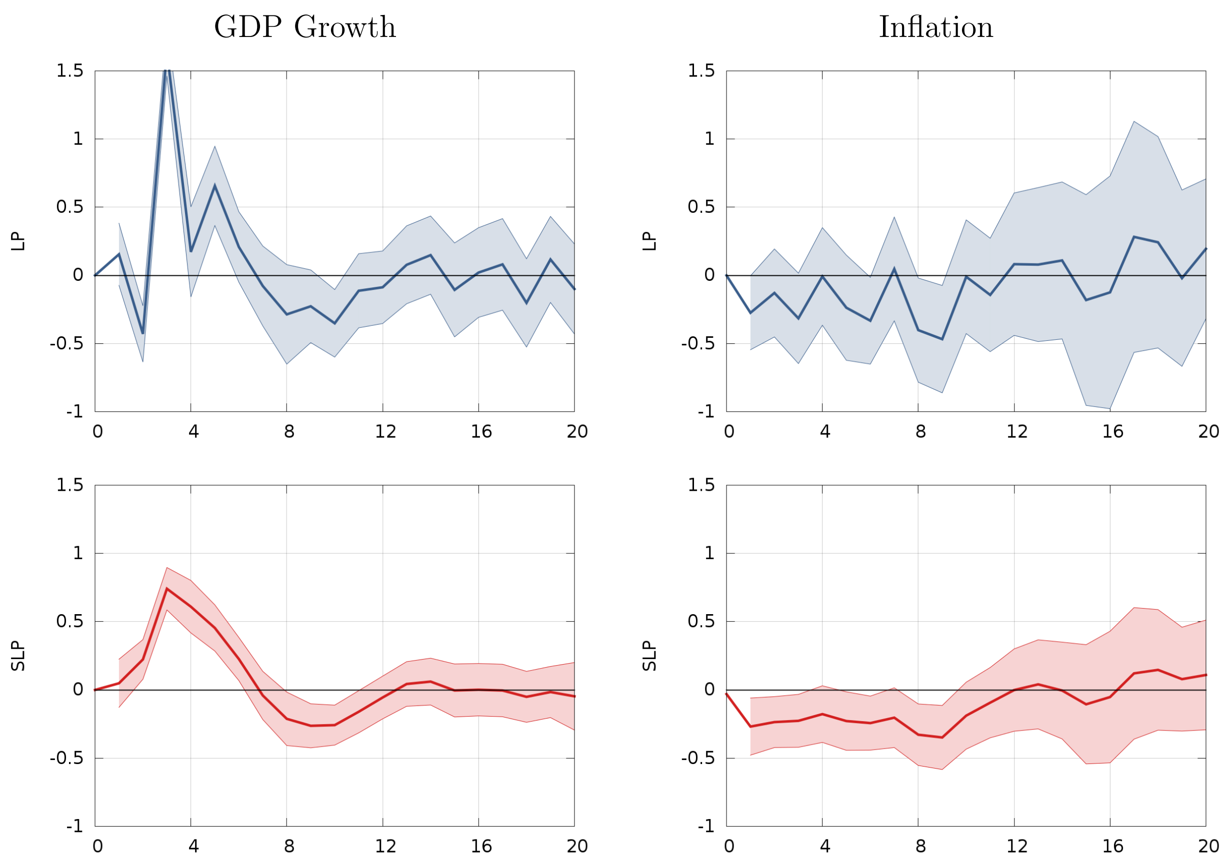
The figure displays the IR of GDP growth (left panels) and inflation (right panels) to a monetary shock identified using instrumental variables and estimated using LP (top panels) and SLP (bottom panels). The shaded area denotes the 90% confidence interval.

Figure 10: STATE-DEPENDENT IR TO A MONETARY SHOCK



The figure displays the state-dependent IR of GDP growth (left panels) and inflation (right panels) to a monetary shock estimated using LP (top panels) and SLP (bottom panels). The state variable s_t respectively takes the values -1 (a recession), 0, and +1 (an expansion) units of $\sigma(s_t)$.

Figure 11: State multiplier of State-dependent IR to a monetary shock



The figure displays the state multiplier of the state-dependent IR of GDP growth (left panels) and inflation (right panels) to a monetary shock estimated using LP (top panels) and SLP (bottom panels). The shaded area denotes the 90% confidence interval.

A Appendix

In this appendix, we show how a partial identification scheme with a timing restriction is a particular case of "identification through controls" and thus can be easily implemented in a LP estimation.

Recursive identification in a linear DGP

Denote $\boldsymbol{\chi}_t$ an $(L \times 1)$ vector generated by a VAR of order 1 (wlog)

$$\boldsymbol{\chi}_t = \boldsymbol{A}\boldsymbol{\chi}_{t-1} + \boldsymbol{C}\boldsymbol{\varepsilon}_t, \tag{10}$$

where $\boldsymbol{C}$ has full rank and ones on the diagonals, and ε_t is the vector of structural innovations with $E\boldsymbol{\varepsilon}_t = \mathbf{0}$ and $E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{'} = \boldsymbol{\Sigma}$ a positive definite diagonal matrix.

It will be convenient to adopt the following conventions for notation:

- Denote $\boldsymbol{\chi}_{\ell,t}$ the ℓth variable of vector $\boldsymbol{\chi}_t$ and denote ${\boldsymbol{\chi}_t}^{<\ell} = (\boldsymbol{\chi}_{1,t}, ..., \boldsymbol{\chi}_{\ell-1,t})'$ the vector of variables ordered before variable $\boldsymbol{\chi}_{\ell,t}$ in $\boldsymbol{\chi}_t$. Similarly, we can define ${\boldsymbol{\chi}_t}^{\leq\ell}$ or ${\boldsymbol{\chi}_t}^{>\ell}$.
- For a matrix $\boldsymbol{C}$ of size $L \times L$ and $(i, j) \in \{1, ..., L\}^2$, denote $\boldsymbol{C}^{<i,<j}$ the $(i-1)\times(j-1)$ submatrix of $\boldsymbol{C}$ made of the first $(i-1)$ rows and $(j-1)$ columns. Similarly, we denote $\boldsymbol{C}^{>i,>j}$ the $(L-i)\times(L-j)$ submatrix of $\boldsymbol{C}$ made of the last $(L-i)$ rows and $(L-j)$ columns. In the same spirit, we denote $\boldsymbol{C}^{i,<j}$ the submatrix of $\boldsymbol{C}$ made of the ith row and the first $(j-1)$ columns. $\boldsymbol{C}^{i,<j}$ is in fact a row vector. A combination of these notations allows us to denote any submatrix of $\boldsymbol{C}$. Finally, denote $\boldsymbol{C}_{ij}$ the ith row jth column element of $\boldsymbol{C}$.

With these notations, we can now state the partial recursive ordering assumption

Assumption 1 (Partial recursive ordering). *The contemporaneous impact matrix* $\mathbf{C}$ *of dimension* $L \times L$ *is of the form*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^{<\ell,<\ell} & \mathbf{0}^{<\ell,\ell} & \mathbf{0}^{<\ell,>\ell} \\ \mathbf{C}^{\ell,<\ell} & 1 & \mathbf{0}^{\ell,>\ell} \\ \mathbf{C}^{>\ell,<\ell} & \mathbf{C}^{>\ell,\ell} & \mathbf{C}^{>\ell,>\ell} \end{bmatrix}.$$

with $\ell \in \{1, .., L\}$, $\mathbf{C}^{<\ell,<\ell}$ *and* $\mathbf{C}^{>\ell,>\ell}$ *matrices of full rank and* $\mathbf{0}$ *denoting a conformable matrix of zeros.*

Assumption 1 states that the shock of interest $\varepsilon_{\ell,t}$, ordered in ℓth position in $\boldsymbol{\varepsilon}_t$, affects the variables ordered from 1 to $\ell - 1$ with a one period lag, and that the first ℓ variables in $\boldsymbol{\chi}_t$ do not react contemporaneously to shocks ordered after $\varepsilon_{\ell,t}$ in $\boldsymbol{\varepsilon}_t$.

We can now state our result:

Proposition 1. *If* $\mathbf{C}$ *satisfies Assumption 1 (the partial recursive ordering assumption), then an estimate of the structural impulse response of* $\boldsymbol{\chi}_{i,t}$ *to a structural shock* δ *to* $\boldsymbol{\chi}_{\ell,t}$

at horizon h, is given by $\beta_{(h)}\delta$ *with* $\beta_{(h)}$ *the coefficient estimated from the smooth local projection*

$$y_{t+h} = \alpha_{(h)} + \beta_{(h)} x_t + \boldsymbol{\gamma}'_{(h)} \boldsymbol{w}_t + u_{(h),t+h}, \quad t = 1, ..., T-H \tag{11}$$

with $y_t = \boldsymbol{\chi}_{i,t}$, $x_t = \boldsymbol{\chi}_{\ell,t}$ *and* $\boldsymbol{w}'_t = (\boldsymbol{\chi}_t^{<\ell}, \boldsymbol{\chi}_{t-1})'$.

Proof. Iterating on (10), we have

$$\boldsymbol{\chi}_{t+h} = \boldsymbol{A}^{h+1}\boldsymbol{\chi}_{t-1} + \boldsymbol{A}^h \mathbf{C}\varepsilon_t + \boldsymbol{A}^{h-1}\mathbf{C}\varepsilon_{t+1} + ... + \mathbf{C}\varepsilon_{t+h} \tag{12}$$

which immediately gives that the impulse response of vector χ_t to shock $\varepsilon_{\ell,t}$ at horizon h is

$$\boldsymbol{\beta}_{(h)} = \boldsymbol{A}^h \mathbf{C}^{\leq L,\ell}$$

i.e., that the vector of impulse responses to shock $\boldsymbol{\varepsilon}_\ell$ is the ℓth column of the matrix $\boldsymbol{A}^h\boldsymbol{C}$.

We now show that the VAR (10) with a partial recursive ordering implies the local projection representation (11) where $E\chi_{\ell,t}u_{(h),t+h} = 0$, which ensures that OLS provides a consistent estimate of $\boldsymbol{\beta}_{(h)}$.

We first write the reduced-form innovation $\boldsymbol{e}_t = \boldsymbol{C}\boldsymbol{\varepsilon}_t$ as the sum of three $L \times 1$ vectors:

$$\boldsymbol{C}\varepsilon_t = \boldsymbol{C}^{\leq L,<\ell}\varepsilon_t^{<\ell} + \boldsymbol{C}^{\leq L,\ell}\varepsilon_{\ell,t} + \boldsymbol{C}^{\leq L,>\ell}\varepsilon_t^{>\ell}$$

where $\boldsymbol{C}^{\leq L,<\ell} = \begin{pmatrix} \boldsymbol{C}^{<\ell,<\ell} \\ \boldsymbol{C}^{\ell,<\ell} \\ \boldsymbol{C}^{>\ell,<\ell} \end{pmatrix}$, $\boldsymbol{C}^{\leq L,\ell} = \begin{pmatrix} \mathbf{0}^{<\ell,\ell} \\ 1 \\ \boldsymbol{C}^{>\ell,\ell} \end{pmatrix}$ and $\boldsymbol{C}^{\leq L,>\ell} = \begin{pmatrix} \mathbf{0}^{<\ell,>\ell} \\ \mathbf{0}^{\ell,>\ell} \\ \boldsymbol{C}^{>\ell,>\ell} \end{pmatrix}$ are the three "block-columns" of $\boldsymbol{C}$ shown in (1). The first component captures the contribution to $\boldsymbol{\chi}_t$ of shocks ordered before $\varepsilon_{\ell,t}$, the second component captures the contribution of $\varepsilon_{\ell,t}$, and the third component captures the contribution of shocks ordered after $\varepsilon_{\ell,t}$.

Thanks to the recursive ordering assumption, $\varepsilon_t^{<\ell}$, the vector of shocks ordered before $\varepsilon_{\ell,t}$, is a function of contemporaneous values of $\boldsymbol{\chi}_t^{<\ell}$ and past values of $\boldsymbol{\chi}_{t-1}$, and can then be recovered from observables (i.e., controls) at time t. Specifically, because the upper-right block of $\boldsymbol{C}$ is filled with zero, we have

$$\begin{aligned} \varepsilon_t^{<\ell} &= \left(\boldsymbol{C}^{<\ell,<\ell}\right)^{-1} e_t^{<\ell} \\ &= \left(\boldsymbol{C}^{<\ell,<\ell}\right)^{-1} (\boldsymbol{\chi}_t - \boldsymbol{A}\boldsymbol{\chi}_{t-1})^{<\ell} \\ &= \left(\boldsymbol{C}^{<\ell,<\ell}\right)^{-1} \boldsymbol{\chi}_t^{<\ell} - \left(\boldsymbol{C}^{<\ell,<\ell}\right)^{-1} (\boldsymbol{A}\boldsymbol{\chi}_{t-1})^{<\ell}. \end{aligned} \tag{13}$$

Similarly, $\varepsilon_{\ell,t}$, the shock of interest, is a function of the contemporaneous value of $\boldsymbol{\chi}_{\ell,t}$, $\boldsymbol{\chi}_t^{<\ell}$ and past values of $\boldsymbol{\chi}_{t-1}$. Indeed, from the VAR, we have

$$\boldsymbol{\chi}_{\ell,t} = (\boldsymbol{A}\boldsymbol{\chi}_{t-1})_\ell + \varepsilon_{\ell,t} + \boldsymbol{C}^{\ell,<\ell}\varepsilon_t^{<\ell},$$

so that using (13) we get

$$\varepsilon_{\ell,t} = \boldsymbol{\chi}_{\ell,t} - \boldsymbol{C}^{\ell,<\ell}\left(\boldsymbol{C}^{<\ell,<\ell}\right)^{-1}\boldsymbol{\chi}_t^{<\ell} - \boldsymbol{C}^{\ell,<\ell}\left(\boldsymbol{C}^{<\ell,<\ell}\right)^{-1}(\boldsymbol{A}\boldsymbol{\chi}_{t-1})^{<\ell} - (\boldsymbol{A}\boldsymbol{\chi}_{t-1})_\ell. \tag{14}$$

Substituting the components of (13) and (14) into (12), we get

$$\begin{aligned} Y_{t+h} &= \boldsymbol{A}^{h+1} Y_{t-1} + \boldsymbol{A}^h \boldsymbol{C} \boldsymbol{\varepsilon}_t + \boldsymbol{A}^{h-1} \boldsymbol{C} \boldsymbol{\varepsilon}_{t+1} + ... + \boldsymbol{C} \boldsymbol{\varepsilon}_{t+h} \\ &= \underbrace{\boldsymbol{A}^h \boldsymbol{C}^{\leq L,\ell}}_{\boldsymbol{\beta}_{(h)}} \boldsymbol{\chi}_{\ell,t} + \boldsymbol{D} \boldsymbol{\chi}_t^{<\ell} + \boldsymbol{B} \boldsymbol{\chi}_{t-1} + \boldsymbol{u}_{t+h} \end{aligned} \tag{15}$$

with $\boldsymbol{D}$ and $\boldsymbol{B}$ some $(L \times L)$ matrices that depend on $\boldsymbol{A}$ and $\boldsymbol{C}$, and where $\boldsymbol{u}_{t+h} = \boldsymbol{A}^h \begin{pmatrix} \mathbf{0}^{<\ell,>\ell} \\ \mathbf{0}^{\ell,>\ell} \\ \boldsymbol{C}^{>\ell,>\ell} \end{pmatrix} \varepsilon_t^{>\ell} + f(\boldsymbol{\varepsilon}_{t+h})_{h>0}$, where $f(.)_{h>0}$ denotes a function of shocks at time $t+h$, $h>0$.

Expression (15) is a Local Projection where the IR vector at horizon h, $\boldsymbol{\beta}_{(h)}$, is given by the coefficient on $\chi_{\ell,t}$. Thanks to the recursive assumption, we have $E\boldsymbol{u}_{t+h}\boldsymbol{\chi}_{\ell,t} = \mathbf{0}$, since $E\varepsilon_t^{>j}\chi_{\ell,t} = 0$ (and since $E\varepsilon_{t+h}\chi_{\ell,t} = 0$), so that OLS on (15) consistently estimates $\boldsymbol{\beta}_{(h)}$, as stated in Proposition 1.

□

Recursive identification in a DGP with state dependence

We now extend Proposition 1 to a DGP with state dependence.

Denote $\boldsymbol{\chi}_t$ an $L \times 1$ vector given by the structural VMA model

$$\boldsymbol{\chi}_t = \sum_{k=0}^{K} \boldsymbol{C}_k(s_{t-k})\boldsymbol{\varepsilon}_{t-k} \tag{16}$$

where $\boldsymbol{\varepsilon}_t$ is the vector of stuctural shocks at time t and $\boldsymbol{C}_j(s_{t-j})$ is the matrix of impulse response functions at horizon k. K is either finite or infinite with $\sum\limits_{k=0}^{\infty} (\boldsymbol{C}_k(s_{t-k}))^2 < \infty$.

With state dependence, the value of the stationary time series s_t affects the impulse responses to a structural shock to $\boldsymbol{\chi}_{\ell,t}$ according to

$$\boldsymbol{C}_j^{\leq L,\ell}(s_{t-k}) = \boldsymbol{\Gamma}_k^{\leq L,\ell} + \boldsymbol{\Theta}_k^{\leq L,\ell} s_{t-k}$$

where $\boldsymbol{\Gamma}_k^{\leq L,\ell}$ and $\boldsymbol{\Theta}_k^{\leq L,\ell}$ are two $(L-\ell)$ column vectors.

Note that model (16) includes a VAR DGP –in fact, a VAR(∞)– as a special case when there is no state dependence.

To allow identification, we posit a partial recursive ordering as in Assumption 1 while allowing for state dependence for the effect of the identified shock, that is by allowing that $\mathbf{C}_\mathbf{0}^{>\ell,\ell}$ to depend on s_t with $\mathbf{C}_\mathbf{0}^{>\ell,\ell}(s_t) = \boldsymbol{\Gamma}_0^{>\ell,\ell} + \boldsymbol{\Theta}_0^{>\ell,\ell} s_t$.

We now show how to obtain the non-linear IR to to a structural shock to $\boldsymbol{\chi}_{\ell,t}$:

Proposition 2. *If $\mathbf{C}_0$ satisfies Assumption 1 with $\mathbf{C}_\mathbf{0}^{>\ell,\ell}(s_t) = \boldsymbol{\Gamma}_0^{>\ell,\ell} + \boldsymbol{\Theta}_0^{>\ell,\ell} s_t$, then an estimate of the structural impulse response of $\boldsymbol{\chi}_{i,t}$ to a structural shock δ to $\boldsymbol{\chi}_{\ell,t}$ at horizon h, is given by $\left(\beta_{0,(h)} + \beta_{1,(h)} s_t\right)\delta$ with $\beta_{0,(h)}$ and $\beta_{1,(h)}$ the coefficients estimated from the smooth local projection*

$$y_{t+h} = \alpha_{(h)} + \beta_{0,(h)} x_t + \beta_{1,(h)} x_t s_t + \boldsymbol{\gamma}'_{(h)} \boldsymbol{w}_t + u_{(h)\,t+h}, \quad t = 1, ..., T-H \tag{17}$$

with $y_t = \boldsymbol{\chi}_{i,t}$, $x_t = \boldsymbol{\chi}_{\ell,t}$ *and* $\boldsymbol{w}_t' = (\boldsymbol{\chi}_{<\ell,t}, \boldsymbol{\chi}_{<\ell,t}s_t, \{\boldsymbol{\chi}_{t-k}\}_{k=1}^{\infty}, \{\boldsymbol{\chi}_{t-k}s_{t-k}\}_{k=1}^{\infty})'$.

Proof. We first show the following Lemma:

LEMMA 1. *There exist* $(L \times L)$ *matrices* P_1, P_2, Q_1, Q_2, R_1, R_2 *(linear functions of* $\{\boldsymbol{C}_k\}_{k=0}^{K}$*) such that the vector of contemporaneous shocks* $\boldsymbol{\varepsilon}_t$ *can be expressed as*

$$\begin{cases} \boldsymbol{\varepsilon}_t^{<\ell} = \left(\mathbf{C}_0^{<\ell,<\ell}\right)^{-1} \boldsymbol{\chi}_t^{<\ell} + \sum_{j=1}^{\infty} (G_1 + G_2 s_t) Y_{t-j} \\ \boldsymbol{\varepsilon}_{\ell,t} = \boldsymbol{\chi}_{\ell t} + \mathbf{C}_0^{\ell,<\ell} \varepsilon_t^{<\ell} + \sum_{j=1}^{\infty} (H_1 + H_2 s_t) Y_{t-j} \\ \boldsymbol{\varepsilon}_t^{>\ell} = \left(\mathbf{C}_0^{>\ell,>\ell}\right)^{-1} \left(\boldsymbol{\chi}_t^{>\ell} - \mathbf{C}_0^{>\ell,\leq\ell}(s_t)\varepsilon_t^{\leq\ell} + \sum_{j=1}^{\infty} (P_1 + P_2 s_t) Y_{t-j}\right) \end{cases} \tag{18}$$

Proof. We proceed by induction and posit that (18) holds for $\{\varepsilon_{t-k}\}_{k>0}$. Starting from the DGP of $\boldsymbol{\chi}_t$ and rearranging to put $\boldsymbol{\varepsilon}_{t-k}$ on the left-hand side, we can express $\{\varepsilon_{t-k}\}_{k>0}$ as a function of lagged values of $\boldsymbol{\chi}_t$ using (18) for $\{t-k\}_{k=1}^{\infty}$ and obtain (18). □

Using (18), we can now express the time t shocks as function of current and past observables Y_t, i.e., re-write the DGP of $\boldsymbol{\chi}_{t+h}$ as a local projection. To see that, first re-write the DGP of $\boldsymbol{\chi}_{t+h}$ as

$$\boldsymbol{\chi}_{t+h} = \boldsymbol{C}_k(s_{t-k})\boldsymbol{\varepsilon}_t + \sum_{k=1}^{\infty} \boldsymbol{C}_k(s_{t-k})\boldsymbol{\varepsilon}_{t-k} + f(\boldsymbol{\varepsilon}_{t+h})_{h>0},$$

where $f(.)_{h>0}$ denotes a function of shocks at time $t+h$, $h>0$.

Then, combining with (18) to express $\boldsymbol{\varepsilon}_t$ as a function of contemporaneous and lagged values of Y_t, we get

$$\boldsymbol{\chi}_{t+h} = \boldsymbol{\Gamma}_k \boldsymbol{\chi}_{\ell,t} + \boldsymbol{\Theta}_k \boldsymbol{\chi}_{\ell,t} s_{t-k} + \mathbf{D}_0 \boldsymbol{\chi}_t^{<\ell} + \mathbf{D}_1 \boldsymbol{\chi}_t^{<\ell} s_t + \sum_{k=1}^{\infty} (\mathbf{B}_0 + \mathbf{B}_1 s_t) \boldsymbol{\chi}_{t-k} + \boldsymbol{u}_{t+h}, \tag{19}$$

with $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{B}_0$ and $\mathbf{B}_1$ some $(L \times L)$ matrices that are linear functions of $\{\boldsymbol{C}_k\}_{k=0}^{K}$ and where $\boldsymbol{u}_{t+h} = \mathbf{F} \begin{pmatrix} \mathbf{0}^{<\ell,>\ell} \\ \mathbf{0}^{j,>\ell} \\ \mathbf{C}_0^{>\ell,>\ell} \end{pmatrix} \boldsymbol{\varepsilon}_t^{>\ell} + f(\boldsymbol{\varepsilon}_{t+h})_{h>0}$.

Expression (19) is a Local Projection where the IR coefficients are given by the coefficient of $\boldsymbol{\chi}_{\ell,t}$ and $\boldsymbol{\chi}_{\ell,t}s_t$. Thanks to the recursive assumption, we have $Eu_{t+h}\boldsymbol{\chi}_{\ell,t} = 0$, since $E\varepsilon_t^{>\ell}\chi_{\ell,t} = 0$ (and since $E\varepsilon_{t+h}\chi_{\ell,t} = 0$), so that OLS on (19) consistently estimates $\boldsymbol{\Gamma}_k$ and $\boldsymbol{\Theta}_k$, as stated in Proposition 2.

□