DYNAMIC EFFECTS OF PERSISTENT SHOCKS

2019

Mario Alloza, Jesús Gonzalo and Carlos Sanz

Documentos de Trabajo
N.º 1944

BANCO DE ESPAÑA
Eurosistema

DYNAMIC EFFECTS OF PERSISTENT SHOCKS (*)

Mario Alloza and Carlos Sanz

BANCO DE ESPAÑA

Jesús Gonzalo

UNIVERSIDAD CARLOS III DE MADRID

(*) We thank Fabio Canova, Jesús Fernández-Villaverde, Alessandro Galesi, Gergely Gánics, Juan F. Jimeno, Mikkel Plagborg-Møller, Juan Rubio-Ramírez, Enrique Sentana, and seminar participants at the I Workshop of the Spanish Macroeconomics Network (Universidad Pública de Navarra), Banco de España, CFE 2018 (University of Pisa), II Workshop in Structural VAR models (Queen Mary University of London), VII Workshop on Empirical Macroeconomics (Ghent University), 2019 American Meeting of the Econometric Society (University of Washington), and 2019 edition of the Padova Macro Talks for insightful comments.
Alloza: m.alloza@bde.es. Gonzalo: jesus.gonzalo@uc3m.es. Sanz: carlossanz@bde.es.

Documentos de Trabajo. N.º 1944

2019

ISSN: 1579-8666 (on line)

Abstract

We show that several shocks identified without restrictions from a model, and frequently used in the empirical literature, display some persistence. We demonstrate that the two leading methods to recover impulse responses to shocks (moving average representations and local projections) treat persistence differently, hence identifying different objects. In particular, standard local projections identify responses that include an effect due to the persistence of the shock, while moving average representations implicitly account for it. We propose methods to re-establish the equivalence between local projections and moving average representations. In particular, the inclusion of leads of the shock in local projections allows to control for its persistence and renders the resulting responses equivalent to those associated to counterfactual non-serially correlated shocks. We apply this method to well-known empirical work on fiscal and monetary policy and find that accounting for persistence has a sizable impact on the estimates of dynamic effects.

Keywords: impulse response function, local projection, shock, fiscal policy, monetary policy.

JEL classification: C32, E32, E52, E62.

Resumen

En este documento mostramos que varios *shocks* identificados sin restricciones de un modelo, y usados frecuentemente en la literatura empírica, son persistentes. Demostramos que los dos principales métodos para recuperar respuestas a impulsos de *shocks* (representaciones de media móvil y proyecciones locales) tratan la persistencia de distinta manera y, por tanto, identifican objetos diferentes. En particular, las proyecciones locales estándar identifican respuestas que incluyen un efecto debido a la persistencia del *shock,* mientras que las representaciones de media móvil implícitamente controlan por la persistencia. Proponemos métodos para restablecer la equivalencia entre proyecciones locales y representaciones de media móvil. En particular, la inclusión de adelantos del *shock* en las proyecciones locales permite controlar por su persistencia y hace que las respuestas resultantes sean equivalentes a las asociadas a *shocks* contrafactuales no autocorrelacionados. Aplicamos este método a trabajos empíricos sobre política fiscal y monetaria y encontramos que controlar por la persistencia tiene un impacto considerable sobre las estimaciones de efectos dinámicos.

Palabras clave: función de respuesta a un impulso, proyecciones locales, *shock,* política fiscal, política monetaria.

Códigos JEL: C32, E32, E52, E62.

1 Introduction

Understanding the origin and propagation of economic shocks has been an important yet elusive challenge in macroeconomics. Early work on this subject has traditionally relied on systems of equations coupled with restrictions implied by economic theory in order to identify economically meaningful shocks. In recent years, researchers have sought the identification of shocks without the use of an empirical model, known as narrative identification, typically by looking at written official documentation, periodicals, magazines, etc. and exploiting arguably exogenous variation in these series. As argued by Ramey (2016), these shocks are meant to be the empirical counterparts of the structural shocks found in theoretical models and hence they should be uncorrelated with both current and lagged values of endogenous variables, and with other shocks in the system. Furthermore, they should be serially uncorrelated since they are expected to represent unanticipated variation in an exogenous variable.

In this paper, we study the estimation of impulse responses in the presence of persistence in the shock. Our analysis highlights the econometric implications of relaxing the assumption of serial uncorrelation in shocks, and quantifies its importance by revisiting well-known empirical applications.

We begin by providing evidence that many shocks used by the literature display serial correlation. These aggregate shocks have been used to estimate the impact of exogenous changes in tax revenues, government spending, and monetary policy on output and other macroeconomic variables. To interpret the dynamic responses to these shocks, it is usually assumed that they are serially uncorrelated or that the researcher can effectively account for this feature.

Next, we analyze the econometric consequences of using persistent shocks in empirical work. When a shock is identified from outside an empirical model, the researcher can recover its dynamic effects (the impulse responses) using either local projections (LPs), as proposed by Jordà (2005), or moving average (MA) regressions.[1] We show that, if the shock is serially uncorrelated, both methods recover the same impulse response functions. However, this equivalence breaks down upon the presence of persistence. In this case, LPs and MA regressions identify different dynamic effects, since they allow for the presence of different channels of propagation when constructing impulse responses.

When persistence exists but is ignored, LPs estimate dynamic effects that contain two components: an economic effect (the economic impact of the shock on the endogenous variables) and an effect that exclusively depends on the degree of serial correlation of the shock. The intuition follows from the way that LPs compute the response at horizon h, regressing the outcome variable in $t+h$ against the shock in time t. Since the standard setting does

[1]By MA regressions, we refer to either single-equation distributed-lags models of an outcome variable against the contemporaneous value and lags of the shock, as in Romer and Romer (2010) (often known as truncated MA), or to a MA structure of the shock embedded in a vector autoregressions (VAR), as in Mertens and Ravn (2012).

not account for how the shock evolves between t and $t+h$, the responses include this persistence in the shock. Perhaps contrary to intuition, including lags of the shock in LPs fails to isolate both effects.[2] Instead, we show that the inclusion of *leads* of the shock in LPs does control for the presence of persistence and deliver responses to the shock *as if* it were serially uncorrelated. Intuitively, the leads control for the evolution of the shock between t and $t+h$. Following this intuition, for some data generating processes, the persistence of a shock may still affect the impulse responses even if the shock is used as an instrument for an endogenous variable in LPs (known as LP-IV; see Stock and Watson (2018) and Ramey and Zubairy (2018)). In this case, the inclusion of leads of the instrument as explanatory variables allows to identify the dynamic effects as if the instrument did not have persistence.[3]

The persistence of a shock does not affect the impulse responses when these are constructed using MA regressions (e.g. using a VAR with the shock as an exogenous variable). In this case, regardless of the existence of persistence, responses always resemble those as if the shock had no persistence, as long as the (truncated) MA representation includes a sufficiently long lag structure. An important consequence of this result is that, even under comparable conditions (e.g., same sample length, same control variables), MA and LPs methods will yield different dynamic estimates in the presence of persistence.[4]

Our results should not be interpreted as a support for one particular method (LPs *vs.* MA regressions) when the process is known to be linear. Instead, they are intended to raise awareness on the identification challenges that arise when shocks follow more complicated data generating processes. The choice of a particular method should consider two issues: (i) the object that is aimed to be identified (a response that includes or not the effect of persistence), and (ii) the efficiency-robustness trade-off of the estimation method (with LPs being more robust but less efficient). Note that our results show that (i) should not be binding, since our proposed methods can adapt LPs to identify the same object as MA regressions, and *viceversa.*

In the last section, we explore how estimates of government spending multipliers, fiscal consolidations, and other monetary policy shocks change when persistence is accounted for. We find that adjusting the estimates for serial correlation has a sizable impact.

In particular, we show that the response of government spending and output to an exogenous shock (measured as news about future changes in defense spending as in Ramey (2011)) is up to 70% higher due to the persistence of the shock. We also contribute to the ongoing debate on the state-dependence of fiscal multipliers by showing that this persistence affects the dynamics of the government spending multiplier during times of expansion and slack.

[2]Another possibility is to include the identified shock as and endogenous variable in a VAR, but note that, as highlighted by Plagborg-Møller and Wolf (2019), this would identify the same object as LPs. We illustrate this point in the appendix.

[3]In particular, we show that the inclusion of leads of the instrument can potentially re-establish the condition of lead-lag exogeneity postulated in Stock and Watson (2018).

[4]We also show how to obtain impulse responses estimated by MA regressions that incorporate the effect of serial correlation.

Ramey and Zubairy (2018) find that, in recessions, the multiplier becomes substantially *negative* immediately after the shock (with estimates close to -2) and that, after two years, it is not different from the multiplier in expansions.[5] We show that accounting for persistence changes the dynamics of the fiscal multiplier in recessions: it becomes smaller upon impact (-1 instead of -2) and remains lower than the multiplier in expansions even after two years.

We also show that ignoring the presence of persistence when using LPs has substantial effects in other contexts. We find that the large negative effects that fiscal consolidations have on output in Guajardo et al. (2014) are partially due to the presence of serial correlation in the shocks. When accounting for their persistence, the negative impact of consolidations on economic activity becomes non-significant during much of the response horizon. We also explore other empirical applications and find that the effects of monetary policy on economic activity can be over or under-estimated depending on the nature of the persistence found in common measures of monetary policy shocks, such as Romer and Romer (2004) and Gertler and Karadi (2015). Additionally, we find that when a shock does not display persistence (as is the case with the measure of tax changes from Romer and Romer (2010)), the (unnecessary) inclusion of leads in LPs does not affect the point estimation of dynamic responses.

In this paper, we focus on shocks, which are generally thought to be serially uncorrelated variables. However, the inclusion of leads of a variable in LPs has more general applications. In particular, a researcher interested in using LPs to uncover the dynamic relations of two variables may be interested in including leads of a third variable to construct counterfactual responses as if the behavior of that third variable had remained constant over the response horizon.[6]

Related literature. Our results relate to several strands of the literature. First, we contribute to previous work that develops the econometric methods of LPs and estimation of dynamic effects. The use of LPs to estimate impulse responses was proposed in the seminal paper of Jordà (2005), which shows the advantages of using LPs over VAR methods, particularly their robustness to certain sources of misspecification and the possibility to accommodate non-linear estimations in a practical way. Since LPs impose fewer specification restrictions than a VAR, the estimates are often more volatile than with MA regressions. Barnichon and Brownlees (2018) propose a strategy, called smooth LPs, that models the sequence of the impulse response coefficients as a linear combination of B-splines. More recently, Plagborg-Møller and Wolf (2019) prove that LPs and VAR methods identify the same impulse responses when both methods have an unrestricted lag structure. This result

[5]See Auerbach and Gorodnichenko (2012) for early work on this topic. The authors find that the government spending multiplier during recessions is higher than during booms.

[6]This can be seen as the counterpart in LPs of constructing counterfactual responses in a VAR that allow to separate a direct effect of a regressor on a dependent variable from other indirect effects. This procedure has been frequently used in the empirical VAR literature. See, for example, Bernanke et al. (1997), Sims and Zha (2006), or Bachmann and Sims (2012).

formalizes some of the examples provided in Ramey (2016), which implies that different identification schemes in a VAR setting can be implemented in a LPs context. Note that our result builds on a different premise: we consider the cases where the shock has already been identified (e.g. using narrative measures) and the researcher wants to use LPs or MA regressions (perhaps embedded in a VAR) to estimate dynamic effects. Our contribution to this literature is to show that, while both LPs and MA regressions identify the same object if the shock is serially uncorrelated, this equivalence breaks down in the presence of persistence.

Our second contribution is to propose a method to re-establish the LP-MA equivalence when there is persistence, namely, the inclusion of leads of the shock in LPs. The use of leads has a long tradition in econometrics, dating back to work on factor analysis by Geweke and Singleton (1981) and on the DOLS estimation of cointegration vectors (Stock and Watson (1993)). Faust and Wright (2011) find that including ex-post forecast errors results in an accuracy improvement when forecasting excess bond and equity returns. More recently, Teulings and Zubanov (2014) find that estimating dynamic effects of a dummy variable (e.g. banking crisis) in a panel data context with fixed effects and LPs suffers from a negative small-sample bias, since the estimation of the fixed effect picks up the value of future realization of the dummy variable. The authors show that this bias is attenuated either by increasing the sample size or by including future realizations of the dummy variable over the response horizon. By contrast, the difference between LPs and MA regressions that we identify is not due to a bias in the estimates, but instead to differences in the identification due to the persistence of the shock. Since our problem still persists asymptotically, increasing the sample does reduce the LP-MA difference. Additionally, this difference is not necessarily negative, but will depend on the nature of the data generating process that drives the persistence.

Third, our paper relates to the literature that employs narrative methods to identify exogenous shocks as an alternative to imposing restrictions in empirical models (e.g. structural VARs). The use of narrative methods dates back to Romer and Romer (1989) and Ramey and Shapiro (1998) and is based on using external information, such as official reports or newspapers, to construct series of shocks that are arguably exogenous to macroeconomic events.[7] Given that the objective of this literature is to identify shocks that have similar properties as those found in theoretical models, persistence is seen as an undesirable feature (Ramey (2016)). Our contribution to this literature is to formally and systematically test for the presence of serial correlation in shocks used by previous empirical work. Although the issue of persistence in shocks has been noted before,[8] we believe ours is the first study to formally test the hypothesis of serial uncorrelation in a variety of narratively-identified shocks ranging from fiscal to monetary policy.

[7]Examples of state of the art narrative methods can be found in Romer and Romer (2004), Romer and Romer (2010), or Ramey (2011).

[8]Ramey (2016) finds that the time aggregation required to convert the shock in Gertler and Karadi (2015) to monthly frequency, inserts serial correlation. Miranda-Agrippino and Ricco (2018) corroborate this finding, by regressing the shock on four lags and testing their joint significance. They also find that other measures of monetary shocks such as Romer and Romer (2004) exhibit serial correlation.

Fourth, our work also relates to empirical applications that use LPs to investigate the dynamic effects of narratively-identified shocks.[9] Our contribution is to show that controlling for persistence in the shocks has substantial impact on the estimated impulse responses when using LPs.

The rest of the paper proceeds as follows. Section 2 provides evidence on the existence of serial correlation in shocks used by prominent previous work. Section 3 describes how LPs and MA regressions treat persistence differently, and proposes a solution to re-establish the equivalence between them. It also provides simulations to help understand the results. Section 4 provides evidence of how serial correlation affects estimates of the effects of fiscal and monetary policies. Section 5 concludes.

2 Evidence of persistence in shocks

Throughout this paper, we consider economic shocks in the same meaning as stated in Ramey (2016), that is, shocks are considered to be the empirical counterpart of those employed in theoretical models. According to this description, empirical shocks should be exogenous to current and lagged endogenous variables, uncorrelated to other exogenous shocks, and represent unanticipated movements (or news about future shocks).

When shocks are identified from within an empirical model, the researcher imposes a set of restrictions to recover shocks that can be economically meaningful, as described above. A byproduct of the restrictions from these models is that the resulting shocks are well-behaved and display desirable statistical features, in particular, no serial correlation.

Alternatively, shocks may be identified without the explicit use of a model. This is typically the case of *narrative* methods. This alternative identification relies on the existence of historical sources (such as official documentation, periodicals, etc...) that allow the researcher to trace the cause and size of such shocks. These methods, which depend crucially on the richness of the data and the judgment of the researcher, offer an excellent opportunity to find exogenous variation in aggregate data. However, due to a lack of a subjacent model, the resulting time series may display undesirable features.

In this section, we provide evidence of persistence in eight prominent aggregate shocks related to monetary and fiscal policy. Some of these shocks are identified using narrative methods, while some employ alternative strategies such as timing restrictions using high-frequency methods. In particular, Romer and Romer (2010) and Cloyne (2013) construct measures of exogenous tax changes for the US and the UK, respectively. The authors classify legislated tax measures according to the motivation, as reflected in official documentation, and consider those tax changes that are the result of causes non-related to the state of the

[9]There are several examples of studies that employ narratively identified shocks to estimate dynamics effects using LPs. To name a few: Owyang et al. (2013) and Ramey and Zubairy (2018) look at the the effects of government spending, Fieldhouse et al. (2017) explore the effects of government assets purchases, and Tenreyro and Thwaites (2016) study the impact of monetary policy using and updated sample from Romer and Romer (2004).

economy. In a similar vein, Ramey and Zubairy (2018) construct a measure of government spending shocks by looking at the announcements of future changes in defense spending. Guajardo et al. (2014) construct a series of fiscal consolidations in OECD countries motivated by a desire to reduce the deficit (as opposed to motivated by current or prospective economic conditions). Romer and Romer (2004) and Cloyne and Hürtgen (2016) identify exogenous changes in monetary policy by looking at the minutes and discussion of the monetary policy committees of the Federal Reserve and Bank of England, respectively (they also orthogonalized the resulting series using forecastable information available at that time). Alternatively, Gertler and Karadi (2015) identify a proxy of monetary policy shocks using high frequency surprises around policy announcements. Lastly, Arezki et al. (2017) construct a measure of news shocks based on the date and size of worldwide giant oil discoveries. While some of these papers employ auxiliary regressions to isolate forecastable information, all have in common that the shocks have not been exclusively identified from a time series model.

To test for the presence of persistence we use a *portmanteau*-type test following Box and Pierce (1970).[10] The null hypothesis is that the data are not serially correlated. We test for the presence of autocorrelation in 40 periods, although results are robust to different horizons (see Table D.1).

The results from these tests are displayed in Table 1. Out of the eight considered shocks, six show very large test statistics that result in rejections of the hypothesis of serial uncor-

Table 1: Persistence in macroeconomic shocks

paper	type of shock	Box-Pierce (40) test	p-value
Arezki et al. (2017)	news about oil discoveries	177.903	0.000
Cloyne (2013)	tax (UK)	98.751	0.000
Cloyne and Hürtgen (2016)	monetary policy (UK)	84.422	0.000
Gertler and Karadi (2015)	monetary policy (US)	124.568	0.000
Guajardo et al. (2014)	fiscal consolidations	185.810	0.000
Ramey and Zubairy (2018)	government spending	182.950	0.000
Romer and Romer (2004)	monetary policy (US)	53.758	0.072
Romer and Romer (2010)	tax (US)	19.023	0.998

The third column implements the Box and Pierce (1970) test of serial correlation using the small sample correction following Ljung and Box (1978). The null hypothesis of this test assumes that the data are not serially correlated within 40 periods. For Arezki et al. (2017) and Guajardo et al. (2014), which refer to panel data, we use a generalized version of the autocorrelation test proposed by Arellano and Bond (1991). The serial correlation test yields p-values smaller than 0.05 when testing the shocks of Romer and Romer (2004) with fewer lags or when using the updated data from Coibion (2012) (p-value drops to 0.0041). Ramey and Zubairy (2018) uses extended data from Ramey (2011).

[10] We implement the small sample correction following Ljung and Box (1978). For the cases of Arezki et al. (2017) and Guajardo et al. (2014), which refer to panel data, we test serial correlation using a generalized version of the autocorrelation test proposed by Arellano and Bond (1991) that specifies the null hypothesis of no autocorrelation at a given lag order.

relation for any level of significance. One of them (Romer and Romer (2004)) displays some degree of serial correlation which leads to failure to reject the null hypothesis only for significance levels above 5%.[11] As further evidence of the presence of serial correlation in the above series, Figure D1 plots the associated correlograms. Romer and Romer (2010) constitutes the only considered shock for which we fail to detect the presence of persistence.

Persistence may have different origins. In some instances, it arises because of the method used to convert a nominal series into real terms. For example, Cloyne (2013) and Arezki et al. (2017) divide their series by lagged GDP, while Ramey and Zubairy (2018) use the GDP deflator and a measure of trend GDP. In other instances, the serial correlation arises because of the mapping between different time frequencies. This is usually the case with the identification of monetary policy shocks, such as Romer and Romer (2004), Gertler and Karadi (2015), or Cloyne and Hürtgen (2016), where daily monetary changes are converted into monthly series. Finally, there are other shocks that are more likely to appear together, because of their multi-period nature (for example, episodes of fiscal consolidations, as identified by Guajardo et al. (2014), tend to be spread over the course a few years) or because of they cluster aorund events like wars (as in Ramey and Zubairy (2018)).

From the econometric point of view, shocks can be seen as forecasting errors. If the loss function is quadratic, the best forecast is the conditional expectation and forecast errors become martingale difference sequence (m.d.s). For other loss functions (e.g. the check function) the forecast errors may not be m.d.s and therefore they can show certain serial correlation.

In light of this evidence, we conclude that many relevant narratively-identified empirical shocks display persistence and hence do not meet one of the characteristics of shocks mentioned above (Ramey (2016)). Hence, we distinguish between strictly-defined shocks, which share all the defining characteristics (including no autocorrelation) and weakly-defined shocks, which may be autocorrelated. We will see that persistent, narratively-identified shocks still contain important variation that allows the researcher to trace the dynamic response of relevant macroeconomic variables. For this reason, in the next section we study how persistence affects the computation of dynamic responses, and how it can be accounted for to allow identification of the objects of interest.

3 Theoretical framework

Without loss of generality, consider the following simple data generating process:

$$\begin{aligned} y_t &= \delta x_t + u_t \\ x_t &= \gamma x_{t-1} + \varepsilon_t, \end{aligned} \tag{1}$$

[11] The hypothesis of serial uncorrelation is rejected for significance levels below 5% when considering fewer lags in the test or when considering a longer series (with updated data) from Coibion (2012). The presence of some degree of autocorrelation is shown in Panel E of Figure D1.

where y_t is the economic outcome variable (for example, GDP), x_t is an economic shock (e.g. a fiscal or monetary policy shock) with $\mathbb{E}(x_t u_t) = 0$, and u_t and ε_t are white noises with mean and variance given by $u_t \sim (\mu_u, \sigma_u^2)$ and $\varepsilon_t \sim (\mu_\varepsilon, \sigma_\varepsilon^2)$. Following the evidence found in the previous section, system (1) allows for the shock variable to be serially correlated. This persistence is captured by the parameter γ.[12] δ measures the contemporaneous impact of variable x_t on y_t and is the main parameter of interest.

The data generating process described by system (1) is intentionally simple to illustrate how the dynamic relationship between the dependent variable y_t and the shock x_t depends on the persistence of the latter. Importantly, the obtained results also arise in more complex settings.[13]

We are interested in recovering the response of our variable of interest when a shock hits the system in period t. This statistic is known as the impulse response function, which we denote by $\mathcal{R}(h)$ for period h:

$$\mathcal{R}(h) = \mathbb{E}\left[y_{t+h} | x_t = 1, \Omega_{t-1}\right] - \mathbb{E}\left[y_{t+h} | x_t = 0, \Omega_{t-1}\right], \tag{2}$$

where Ω_{t-1} represents all the history of previous realizations of ε_t and x_t up to period $t-1$. Importantly, note that the above definition does not condition for future realizations of x_t. Hence, if $\gamma \neq 0$, an initial unit impulse in x_t does not imply that $x_{t+j} = 0$.[14] In other words, equation (2) describes dynamic responses that include the possible persistence of the shock x_t. For example:

$$\begin{aligned}
\mathcal{R}(0) &= \frac{\partial y_t}{\partial x_t} = \delta \\
\mathcal{R}(1) &= \frac{\partial y_{t+1}}{\partial x_t} = \delta\gamma \\
\mathcal{R}(2) &= \frac{\partial y_{t+2}}{\partial x_t} = \delta\gamma^2 \\
& \cdots
\end{aligned}$$

However, the researcher might also be interested in the response to the shock *as if* the shock had no persistence. We call this response $\mathcal{R}(h)^*$ and define it as:

$$\mathcal{R}(h)^* = \mathbb{E}\left[y_{t+h} | x_t = 1, x_{t+1}, ..., x_{t+h}, \Omega_{t-1}\right] - \mathbb{E}\left[y_{t+h} | x_t = 0, x_{t+1}, ..., x_{t+h}, \Omega_{t-1}\right]. \tag{3}$$

Contrary to equation (2), equation (3) controls for future realizations of x_t so that it describes dynamic responses that do not incorporate the effect of persistence (regardless of the value of γ), i.e. the responses are observationally equivalent to those that would arise

[12] Although we think of x_t as a shock (probably resulting from a narrative identification), it could be in principle any macroeconomic variable that displays persistence.

[13] In Subsection 3.3, we consider more complex models that include persistence in the dependent variable, lagged effects of the shock, and x_t as an instrument of a true shock.

[14] This impulse response is equivalent to $\mathcal{R}(h) = \mathbb{E}\left[y_{t+h} | \varepsilon_t = 1, \varepsilon_{t+1} = 0, ..., \varepsilon_{t+h} = 0, \Omega_{t-1}\right] - \mathbb{E}\left[y_{t+h} | \varepsilon_t = 0, \varepsilon_{t+1} = 0, ..., \varepsilon_{t+h} = 0, \Omega_{t-1}\right]$. See, for example, Koop et al. (1996).

from a data generating process with $\gamma = 0$:

$$\begin{aligned}
\mathcal{R}(0)^* &= \frac{\partial y_t}{\partial x_t} = \delta \\
\mathcal{R}(1)^* &= \frac{\partial y_{t+1}}{\partial x_t}\Big|_{x_{t+1}} = 0 \\
\mathcal{R}(2)^* &= \frac{\partial y_{t+2}}{\partial x_t}\Big|_{x_{t+1},x_{t+2}} = 0 \\
& \dots
\end{aligned}$$

Note that, if $\gamma = 0$ (the shock is not persistent), then $\mathcal{R}(h) = \mathcal{R}(h)^* \ \forall \ h$. By contrast, if $\gamma \neq 0$, then $\mathcal{R}(h) \neq \mathcal{R}(h)^* \ \forall \ h > 0$.

3.1 Differences between MA regressions and LPs under persistence

We now consider the two most frequently used methods to estimate impulse responses, MA regressions and LPs, and compare the objects that they identify when the shock is persistent. We first consider the case of MA regressions. This is the underlying method employed by VARs to recover the response to a shock and its use is widespread in applied macroeconomics. In the case of system (1), note that we can recover the response function $\mathcal{R}(h)^{MA}$ using the following regression:[15]

$$y_t = \theta_0 x_t + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \theta_3 x_{t-3} + \theta_4 x_{t-4} + \ldots + e_t, \tag{4}$$

and it follows that $\mathcal{R}(h)^{MA} = \frac{\partial y_{t+h}}{\partial x_t} = \theta_h \ \forall \ h$.

The second main method to compute impulse responses is LPs, proposed by Jordà (2005). LPs are more robust to certain sources of misspecification and for this reason, their use has increased in recent times (see Ramey (2016) for examples). LPs compute impulse responses by estimating an equation for each response horizon $h = 0, 1, \ldots, H$:

$$y_{t+h} = \delta_h x_t + \xi_{t+h}, \tag{5}$$

where the sequence of coefficients $\{\delta_h\}_{h=0}^H$ determine the response of the variable of interest $\mathcal{R}(h)^{LP} = \delta_h$ for each horizon h.[16]

We now consider under which conditions both methods identify the same objects.

Proposition 1. *Given the data generating process described by system* (1)*, if the shock x_t is serially uncorrelated, then the response functions identified by MA regressions and LPs are equal for all response horizons, that is:*

If $\gamma = 0$, then $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^{LP} = \mathcal{R}(h)^* = \mathcal{R}(h) \ \forall h$.

[15] This regression should include as many lags as the response horizon $h = 0, 1, \ldots, H$.

[16] Unrelated to our case at hand, note that the structure of the LPs induce serial correlation in the residuals ξ_{t+h}. This is usually corrected by computing autocorrelation-robust standard errors (Jordà (2005)).

If the shock is serially correlated, then the response functions identified by MA regressions and LPs are different for all $h > 0$:

If $\gamma \neq 0$ and $h = 0$, then $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^{LP} = \mathcal{R}(h)^{*} = \mathcal{R}(h)$.

If $\gamma \neq 0$ and $h \geq 1$, then $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^{*} \neq \mathcal{R}(h)^{LP} = \mathcal{R}(h)$.

Proof. See Appendix A.1. □

Following the above proposition, when $\gamma \neq 0$, LPs recover a dynamic response that includes three dynamic effects: (i) the effect that x_t has directly on y_{t+h} (due to a lagged impact of the shock), (ii) the effect that x_t has through the persistence of y_t, and (iii) the effect that x_t has on y_{t+h} through x_{t+h} (since $cov(x_t, x_{t+h}) \neq 0$ when $\gamma \neq 0$). The first two effects are frequently the objects that the econometrician aims at recovering. They are independent of γ and are shut down in our simple specification of system (1).[17] The last effect (the *persistence* effect of x_t) drives the difference between $\mathcal{R}(h)^{MA}$ and $\mathcal{R}(h)^{LP}$. In particular, $\mathcal{R}(h)^{LP} = \mathcal{R}(h) = \delta\gamma^h$, while $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^{*} = 0$ for all $h \geq 1$.

To understand why LPs, unlike MA regressions, incorporate this third effect due to the persistence of x_t, consider the LPs when $h = 1$:

$$y_{t+1} = \delta_1 x_t + \xi_{t+1}, \tag{6}$$

where $\delta_1 = \mathcal{R}(1)^{LP}$. The direct effect of x_t on y_{t+1} is 0. If x_t had no persistence, then δ_1 would be 0. However, when $\gamma \neq 0$, we can use system (1) to express y_{t+1} as a function of x_t:

$$\begin{aligned} y_{t+1} &= \delta x_{t+1} + u_{t+1} \\ &= \delta\left(\gamma x_t + \varepsilon_{t+1}\right) + u_{t+1} \\ &= \delta\gamma x_t + u^{*}_{t+1}, \end{aligned}$$

where $u^{*}_{t+1} = \delta\varepsilon_{t+1} + u_{t+1}$. This shows that the coefficient δ_1 in equation (6) will also recover the persistence effect of x_t: $\delta_1 = \delta\gamma$.[18] The intuition is that between period t and period $t+1$, x_t affects x_{t+1} when $\gamma \neq 0$. Since x_{t+1} is not a regressor in equation (6), then this effect is absorbed by δ_1.

When impulse responses are identified using MA regressions, the treatment of the persistence of x_t is different. To see it more clearly, consider a version of equation (4) expressed in terms of $t + 1$:

$$y_{t+1} = \theta_0 x_{t+1} + \theta_1 x_t + \theta_2 x_{t-1} + \theta_3 x_{t-2} + \theta_4 x_{t-3} + \ldots + e_{t+1}. \tag{7}$$

As noted earlier, the sequence of coefficients θ_h determines the response function. Consider the response when $h = 1$, i.e. $\mathcal{R}(1)^{MA} = \theta_1$. Note that while we know from system (1) that

[17] We will incorporate them in our simulation exercises in the next subsection.

[18] When augmenting system (1) so that y_t has persistence as $y_t = \rho y_{t-1} + \delta x_t + u_t$, then the LPs coefficient also recovers this additional effect: $\delta_1 = \delta\rho + \delta\gamma$.

$\frac{\partial y_{t+1}}{\partial x_t} = \delta\gamma$, the coefficient recovered by θ_1 is indeed $\left.\frac{\partial y_{t+1}}{\partial x_t}\right|_{x_{t+1}} = 0$. That is, since the MA representation controls for x_{t+1}, the persistence effect of x_t is accounted for.[19]

In other words, MA regressions identify:

$$\mathcal{R}(h)^{MA} = \mathbb{E}\left[y_{t+h}|x_t = 1, \Omega_{t-1}, x_{t+h-1}, ..., x_{t+1}\right] - \mathbb{E}\left[y_{t+h}|x_t = 0, \Omega_{t-1}, x_{t+h-1}, ..., x_{t+1}\right],$$

while LPs identify:

$$\mathcal{R}(h)^{LP} = \mathbb{E}\left[y_{t+h}|\varepsilon_t = 1, \Omega_{t-1}\right] - \mathbb{E}\left[y_{t+h}|\varepsilon_t = 0, \Omega_{t-1}\right].$$

Note that the difference between $\mathcal{R}^{LP}$ and $\mathcal{R}^{MA}$ is positive (negative) when $\gamma > 0$ ($\gamma < 0$). In empirical applications, γ may be positive or negative.[20]

3.2 Reestablishing the equivalence between MA regressions and LPs

In this subsection we lay out two methods that can render the responses from MA regressions and LPs identical, even under the presence of persistence.

3.2.1 Adapting LPs to exclude the effect of serial correlation

A researcher may be interested in recovering responses as if the shock were serially uncorrelated. However, we have shown that $\mathcal{R}^{LP}(h) \neq \mathcal{R}^{MA}(h)$ if $\gamma \neq 0$ and $h \geq 1$.

Two apparent methods to avoid LPs picking up the effect of persistence in x_t are: (i) to include lags in the regression (5), or (ii) to replace x_t with the error term that purges out the persistence:

$$\varepsilon_t = x_t - \gamma x_{t-1}. \tag{8}$$

However, neither of these methods yields $\mathcal{R}^*(h)$. The reason is that replacing x_t with ε_t does not include any further information between t and $t+h$, so the responses of the dependent variable will still be affected by x_{t+h}. This point is further developed in Appendix B.1.

A third potential method to exclude the effect of persistence would be recasting system (1) as a VAR that includes the shock as an endogenous variable. However, since in this case LPs and a VAR would identify the same impulse responses (see Plagborg-Møller and Wolf (2019)) the VAR responses would also include an effect due to the persistence of the shock—we explore this in more detail in Appendix B.2.

[19] In practice, the researcher may compute an impulse response function analytically instead of estimating equation (4), by simulating the path of y_t using the first equation of system (1) and setting $x_t = 1$ in time t and $x_t = 0$ for the rest of the periods. The results would be equivalent to those obtained from estimating equation (4): in the first case the researcher *naively* ignores the existence of persistence and in the second case the MA representation implicitly accounts for the effect of persistence.

[20] For example, γ seems to be positive in Ramey and Zubairy (2018), and negative in Romer and Romer (2004).

Instead, we propose a method based on the inclusion of leads of the persistent shock variable. In particular, given the DGP of equation (1), one should regress:

$$y_{t+h} = \delta_{h,0} x_t + \delta_{h,1} x_{t+1} + \xi_{t+h}, \tag{9}$$

where $\delta_{h,0}$ is the h-horizon response identified by LPs that include leads of the shock x_t, which we denote as $\mathcal{R}^F(h)$.

Proposition 2. *Given the data generating process described by system* (1)*, the response function identified by modified LPs to a shock x_t as described in equation* (9) *is equal to the response as if the shock had no persistence (and to the response obtained from MA regressions as in equation* (4)*), that is:*
$\mathcal{R}(h)^F = \mathcal{R}(h)^* = \mathcal{R}(h)^{MA}\ \forall\ \gamma$ *and* h.

Proof. See Appendix A.2. □

Intuitively, leads of x_t in equation (9) act as controls for the persistence of the shock, so that the parameter $\delta_{h,0}$ reflects the dynamic response to a counterfactual serially-uncorrelated shock, that is, controlling for the effect due to $\frac{\partial x_{t+1}}{\partial x_t} \neq 0$ built in system (1) when $\gamma \neq 0$.

In more general processes, in which the autocorrelation of the shock may be of an order larger than one, the optimal choice of leads can be derived adapting the procedure from Choi and Kurozumi (2012). The most conservative procedure would be to include h leads of the shock in each period h. We will revisit this issue in Section 4, when considering empirical applications.

3.2.2 Adapting MA regressions to include the effect of persistence

As noted earlier, $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^*$ regardless of the value of γ. However, in some instances (we discuss this in the next subsection), the researcher may be interested in the response that includes the effect of persistence ($\mathcal{R}(h)$). In this subsection, we show how to adapt MA regressions to recover these responses. Intuitively, the idea is to compute the impulse responses in system (1) with respect to ε_t instead of x_t.

Consider a recursive substitution of x_t in system (1):

$$y_t = \delta\gamma^t x_0 + \delta \sum_{i=0}^{t} \gamma^i \varepsilon_{t-i} + u_t. \tag{10}$$

The responses of y_t to ε_t, which we denote by $\mathcal{R}(h)^{MA-per}$, can be obtained from the coefficients $\tilde{\theta}_h$ in:

$$y_t = \tilde{\theta}_0 \varepsilon_t + \tilde{\theta}_1 \varepsilon_{t-1} + \tilde{\theta}_2 \varepsilon_{t-2} + \tilde{\theta}_3 \varepsilon_{t-3} + \tilde{\theta}_4 \varepsilon_{t-4} + \ldots + e_t. \tag{11}$$

Proposition 3. *Given the data generating process described by system* (1)*, the response function identified by MA regressions of y_t to the innovation ε_t as described in equation* (11) *is equivalent to the response that includes the effects of persistence (and to the response obtained from LPs as in equation* (5)*):*

$\mathcal{R}(h)^{MA-per} = \mathcal{R}(h) = \mathcal{R}(h)^{LP}\ \forall\ \gamma$ *and* h*.*

Proof. See Appendix A.3. □

Proposition 3 establishes a direct equivalence between the coefficients obtained from equation (11) and those obtained from LPs in equation (5): $\tilde{\theta}_h = \delta_h\ \forall\ h$. The former are also related to the coefficients estimated from the MA representation in terms of x_t, as in equation (4): $\theta_0 = \tilde{\theta}_0 = \delta$, $\theta_1 = \tilde{\theta}_1 - \gamma\tilde{\theta}_0, \ldots, \theta_h = \tilde{\theta}_h - \gamma\tilde{\theta}_{h-1}$. Intuitively, the response of y_{t+1} to x_t has an overall effect of $\delta_1 = \tilde{\theta}_1$, which includes (i) the direct effect of x_t on y_{t+1} (0, in our simple case) and (ii) the effect on y_{t+1} that is due to the persistence in x_t (given by $\gamma\delta$). The *standard* MA estimation from equation (4), since it accounts for the evolution of x_t over the response horizon, is implicitly subtracting the part of the response that is given by the persistence of x_t from the overall effect.

3.3 Examples

In this subsection, we perform stochastic simulations of the asymptotic behavior of the impulse response functions using both LPs and MA regressions. Our goal is twofold. First, to evaluate quantitatively the conclusions reached in the previous subsection using a plausible calibration of the parameters that determine the model. Second, to consider a slightly more complex (and realistic) version of the data generating process that includes richer features frequently present in real empirical applications. In particular, we consider the following process:

$$\begin{aligned} y_t &= \rho y_{t-1} + B_0 x_t + B_1 x_{t-1} + u_t \\ x_t &= \gamma x_{t-1} + \varepsilon_t, \end{aligned} \tag{12}$$

where $\mathbb{E}(\varepsilon_{t-s}u_{t-r}) = 0\ \forall s, r \geq 0$, and u_t and ε_t follow $\mathcal{N}(0,1)$ distributions.[21] We set $B_0 = 1.5$, $B_1 = 1$, $\rho = 0.9$, and $\sigma^2_\varepsilon = \sigma^2_u$.

Compared to system (1), the new DGP described in system (12) includes persistence in the outcome variable through ρ, and allows the shock x_t to have lagged effects on y_t through B_1.[22]

We simulate the system above during 100 million periods and recover the dynamic responses of y_t to the shock x_t using LPs:

$$y_{t+h} = \rho y_{t-1} + \beta_{h,0}x_t + \beta_{h,1}x_{t-1} + \beta_{h,f}x_{t+1} + \xi_{t+h}. \tag{13}$$

[21] When $\gamma = 0$, this model is similar to the (truncated) moving average representations of the effects of tax changes of Romer and Romer (2010), but with fewer lags of the shock. When y_t is a vector, this specification is often referred to as VAR-X o VAR with exogenous variables. See, for example Alesina et al. (2015) or Mertens and Ravn (2012).

[22] We introduce this extra lag of the shock to make explicit the distinction between the effect due to the persistence of the shock and the effect of lagged values of the shock on current outcomes.

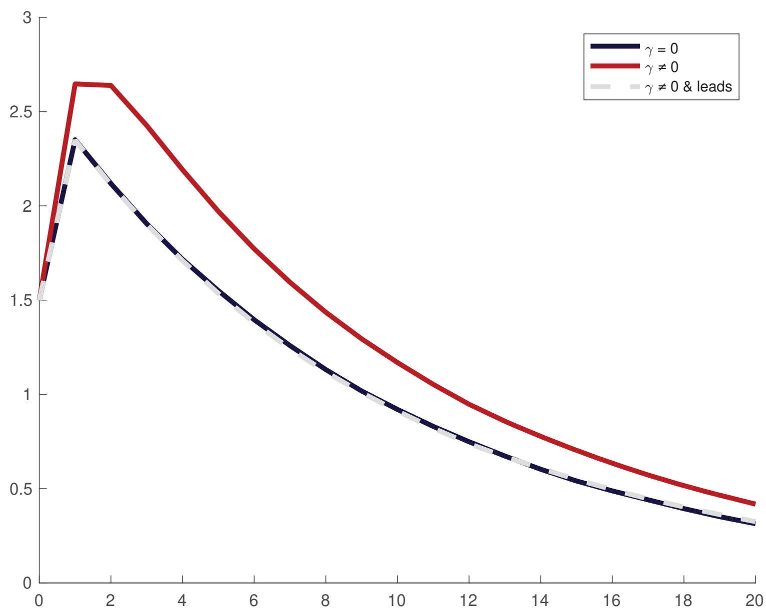
We consider three cases: (i) no persistence ($\gamma = 0$), without including leads in the estimation (i.e., setting $\beta_{h,f} = 0$); (ii) some persistence ($\gamma = 0.2$) and still $\beta_{h,f} = 0$; (iii) some persistence ($\gamma = 0.2$), including a lead of the explanatory variable (i.e. allowing $\beta_{h,f} \neq 0$).[23]

Note that equation (13) must include a lag of shock x_t to effectively capture the effect of B_1 in system (12). However, this does *not* control for the potential persistence of shock x_t, as will be apparent in the simulations.

Figure 1 shows the results of our simulations. In case (i) (dark-blue solid line), the response has a contemporaneous effect of $\hat{\beta}_{1,0} = 1.5$ and peaks at the following period due to the the fact that both ρ and B_1 have positive values. Using the language of the previous section, the impulse response function estimated by LPs with no persistence is asymptotically equivalent to the one obtained directly from equation (13), that is, $\hat{\mathcal{R}}(h)^{LP} \rightarrow \mathcal{R}(h)^*$.

In case (ii) (red solid line), the introduction of persistence in the shock x_t results in a larger effect on y_t on all horizons after impact. This has potentially important implications: if a macroeconomist is interested in the effects of a serially-uncorrelated shock (as in most general equilibrium models), but naively estimates equation (13), implicitly setting $\beta_{h,f} = 0$, then the

Figure 1: Simulated responses using LPs



This figure shows the response of a simulated outcome variable to a shock with different degrees of persistence, using LPs. The dark blue line shows the results of estimating equation (13) assuming $\gamma = 0$ in equation (12). The red line shows the same estimation when $\gamma = 0.2$. The dashed grey line shows the response after including leads of the shock as in equation (13) and still assuming $\gamma = 0.2$.

dynamic response is upwardly biased due to the persistence of the shock, i.e. $\hat{\mathcal{R}}(h)^{LP} > \mathcal{R}(h)^*$ for $h > 0$. Given the assumptions on the autocorrelation of the process x_t, the bias is particularly large in the short and medium run. Higher values of the persistence parameters γ and ρ would increase the difference between both responses (blue and red lines in Figure 1).

[23] The choice of $\gamma = 0.2$ is based on an empirical application that we will present in the next section. Of course, larger values of ρ would yield higher biases due to the persistence of the process.

In case (iii) (dashed grey line in Figure 1), we see that the inclusion of leads of x_t renders the response of the outcome variable to a persistent shock identical to the one obtained when considering a shock without persistence, i.e. $\hat{\mathcal{R}}(h)^F \rightarrow \mathcal{R}(h)^*$. In Appendix B.3 we provide an alternative simulation where the shock x_t in (12) is taken from the actual data.

Next, we use these simulations to show that the computation of impulse responses using MA regressions always yields the same estimates regardless of the persistence in x_t, that is, $\hat{\mathcal{R}}^{MA}(h) \rightarrow \mathcal{R}^*(h)$ for any value of γ.

First, note that, since $\rho < 1$, system (12) can be inverted and re-written as:

$$y_t = (1 - \rho L)^{-1} (B_0 + B_1 L) x_t + (1 - \rho L)^{-1} u_t, \tag{14}$$

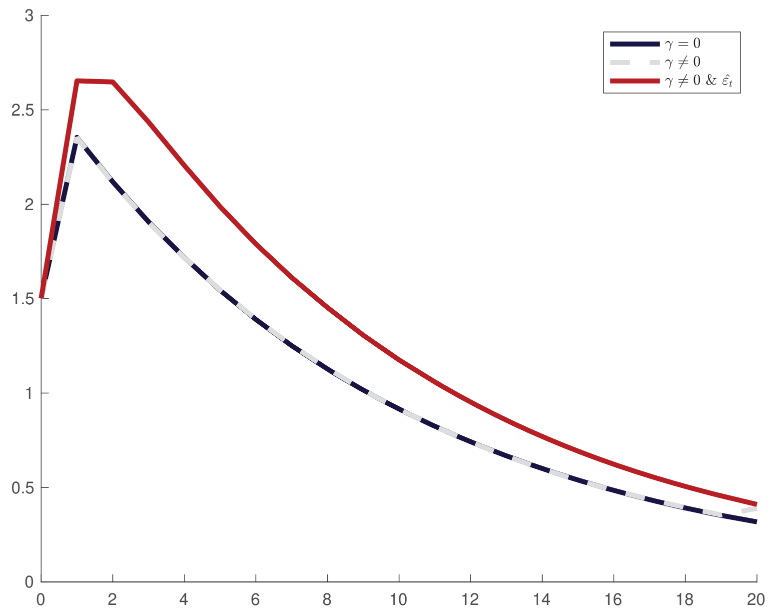
where L represents the lag operator.

Given the independence of u_t and x_t, the representation from equation (14) suggests that the dynamic responses of y_t from x_t can be obtained from the coefficients ϑ_h in the following regression:

$$y_t = \vartheta_0 x_t + \vartheta_1 x_{t-1} + \vartheta_2 x_{t-2} + \vartheta_3 x_{t-3} + \ldots + \vartheta_H x_{t-H} + \xi_t, \tag{15}$$

where H is the response horizon.[24]

Figure 2: Simulated responses using MA regressions



This figure shows the response of a simulated outcome variable to a shock with different degrees of persistence, using MA regressions. The dark blue line shows the results of estimating equation (15) assuming $\gamma = 0$ in equation (12). The dashed grey line shows the same estimation when $\gamma = 0.2$. The red line shows the response when substituting x_t in equation (15) by $\hat{\varepsilon}_t$, an OLS estimate of ε_t (see equation (8)), where serial correlation has been removed.

[24]Baek and Lee (2019) show that for autoregressive distributed lag models, setting the lag order to H is a necessary condition to achieve consistency.

We estimate equation (15) fo three different cases: (i) assuming that $\gamma = 0$ in the data generating process described in system (12), (ii) assuming that $\gamma = 0.2$ and (iii) substituting x_t by $\hat{\varepsilon}_t$ in equation (15) (i.e. following equation (11)).

The results are shown in Figure 2. Cases (i) and (ii) are displayed in blue and dashed grey lines, respectively. As argued earlier, since equation (15) controls for all potential dynamic effects of x_t, including its persistence, the coefficients ϑ_h reflect the responses to a shock as if the variable x_t showed no persistence, regardless of the value of γ. Hence, we have that $\hat{\mathcal{R}}(h)^{MA} \rightarrow \mathcal{R}(h)^*$ for any γ. Note that these impulse response functions are the same as those obtained with LPs ($\hat{\mathcal{R}}(h)^{LP}$) when $\gamma = 0$, or when we include leads in the LPs ($\hat{\mathcal{R}}(h)^F$).

Case (iii) is shown in the red line in Figure 2. As argued in the previous subsection, when computing the impulse response with respect to ε_t, we are allowing the MA regressions to pick up the effect that is due to the persistence in x_t. In other words, since we do not implicitly control for the leads of x_t but for those of ε_t in the MA representation, we are not taking into account the persistence of x_t. In this case, the responses are equal to those obtained from LPs when $\gamma \neq 0$: $\hat{\mathcal{R}}(h)^{MA-per} = \hat{\mathcal{R}}(h)^{LP} \rightarrow \mathcal{R}(h)$.

3.4 Local projections with instrumental variables

Recently, there has been an increased attention to the use of external sources of variation as instruments in LPs or VARs.[25] In this section, we investigate how persistence may affect the estimation of dynamic effects when using instrumental variables in local projections (LP-IV).

Stock and Watson (2018) provide the conditions under which a researcher can exploit external variation to estimate impulse response functions. A valid instrument z_t should be both relevant and contemporaneously exogenous, that is, z_t should not be correlated to any shock in the system except with the one that the researcher is interested in. Lastly, Stock and Watson (2018) impose a restriction called lead/lag exogeneity, which implies that the instrument should not be correlated with any lead or lag of any of the shocks in the system.

Consider the following data generating process:

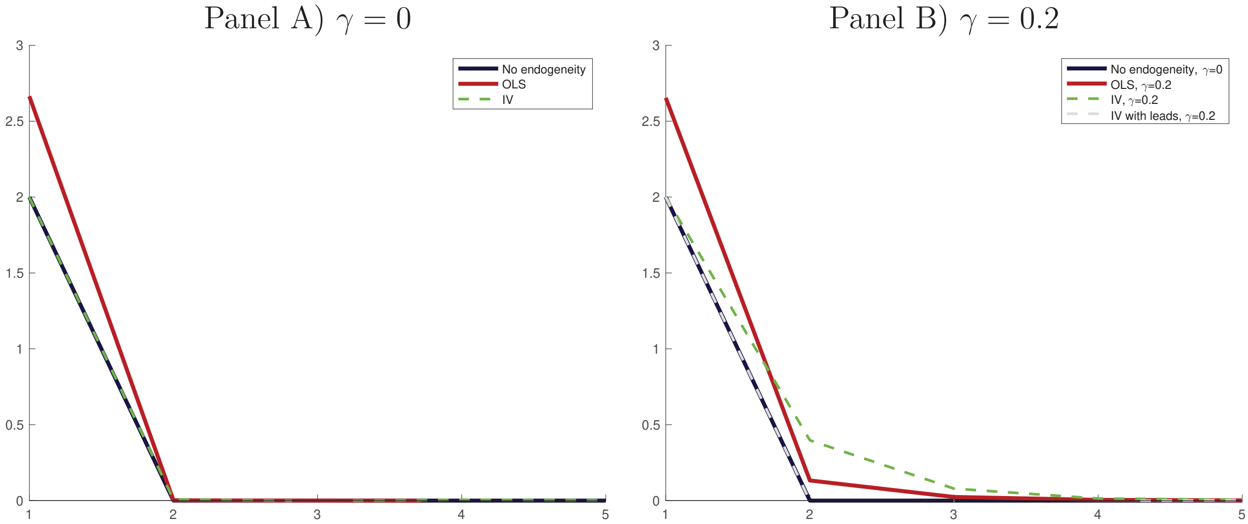
$$\begin{aligned} y_t &= \beta g_t + u_t \\ u_t &= m_t + a_t \\ g_t &= \lambda x_t + (1-\lambda) m_t \\ z_t &= x_t + \nu_t \\ x_t &= \gamma x_{t-1} + \varepsilon_t, \end{aligned} \tag{16}$$

where a_t, ν_t and ε_t follow independent $\mathcal{N}(0,1)$ distributions. A researcher may be interested in estimating the dynamic effects of variable g_t on y_t (e.g. the effects of government

[25] See Mertens and Ravn (2013) for a implementation of this method within a VAR (known as Proxy-SVAR) or Ramey and Zubairy (2018) for an example in a context of LP. Related to this, Ramey (2016) discusses the distinction between shock, innovation, and instrument.

spending on output). However, g_t is endogenous due to the presence of an omitted variable m_t. The researcher may have the availability of an instrument z_t, which is contemporaneously exogenous by construction and relevant when $\lambda \neq 0$. This instrument, since it depends directly on the shock x_t, displays persistence when $\gamma \neq 0$. When there is persistence in the instrument (and the shock), the lead/lag exogeneity condition mentioned above is not satisfied. To illustrate this point, we simulate system (16) setting $\beta = 2$ (and different values of λ and γ) for 100 million periods, and estimate the dynamic effects of g_t on y_t using LP.

Figure 3: LPs with instrumental variables



This figure shows the response of a simulated outcome variable to a shock using local projections with instruments, with an underlying DGP given by system (16) and calibrated for different degrees of persistence in the shock ($\gamma = 0$ in panel A and $\gamma = 0.2$ in Panel B). In both panels, red lines refers to estimation using LPs estimated using OLS and green dashed lines refer to LPs estimated using instrumental variables, when the DGP generates endogeneity. For reference, the blue solid lines (in both panels) display responses when the DGP does not generate persistence or endogeneity. In Panel B, the grey pointed line displays responses estimated using instrumental variables in LPs and including leads of the shock.

We first consider the case of $\gamma = 0$ and $\lambda = 1$, that is, there is no problem of endogeneity or persistence. The estimated effect of g_t on y_t recovered by LPs is represented by a solid blue line in Panel A of Figure 3. As expected, the contemporaneous impact of government spending on output is equal to 2. When considering $\lambda = 0.5$ (but still no persistence, i.e. $\gamma = 0$), LPs that employ OLS will deliver biased estimates of the contemporaneous effect of g_t (solid red line). The difference between the red and the blue lines in the first period is a measure of the endogeneity bias. The problem of endogeneity can be addressed by using z_t as an instrument for g_t to recover the exogenous variation in government spending (given by x_t). This result (still considering $\gamma = 0$) is represented by the dashed grey line in Panel A of Figure 3, which shows how the use of LP-IV can overcome the presence of endogeneity, delivering a response function identical to the benchmark case without omitted variables bias.

Next, we repeat the previous exercise but now we allow for persistence in the instrument (due to persistence of the shock); in particular, we set $\gamma = 0.2$. The results are shown in

Panel B of Figure 3 (we still represent, in solid blue line, the benchmark case of $\gamma = \lambda = 1$ for reference). When there is endogeneity and persistence, LPs estimates of the dynamic effects of g_t are affected by both an endogeneity bias on impact, and by the effect of persistence in the instrument during the rest of the response horizon (as shown in the previous section). This result is displayed by the solid red line in Panel B of Figure 3, which is different from zero after impact. Now consider estimating the dynamic effects using LP-IV with instrument z_t (that displays persistence). The results (dashed green line) show that the use of the instrument addresses the problem of endogeneity (on impact, the effect from the LP-IV estimates is able to recover the true effect of $\beta = 2$). However, the dynamic effect from the rest of the response horizon still reflects the presence of persistence.

As discussed above, persistence in the instrument violates the lead/lag exogeneity condition. Stock and Watson (2018) state that, in general, this condition could be satisfied by the inclusion of further controls in the LP-IV regression. If the source of persistence is strictly restricted to the instrument, Stock and Watson (2018) show that the lead/lag exogeneity condition could be reestablished by including lags of the instrument. However, in cases like system (16), where the instrument inherits its persistence from the shock, lags of the instrument will not satisfy the lead-lag exogeneity condition. We build on intuition laid out by Stock and Watson (2018) and adapt it to the problem of persistence by including leads of the instrument in the set of exogenous variables in the LP-IV estimates. The results, shown in dashed grey lines in Panel B of Figure 3, corroborate this intuition: despite the presence of both endogeneity and persistence, enhancing the LP-IV estimates with leads of the shock allows to recover the dynamic effects *as if* the instrument were not serially correlated.

In sum, the presence of persistence can potentially violate the lead-lag exogeniety assumption, invalidating inference under LP-IV. The solution to reestablish this condition will depend on the source of persistence in the model. When the instrument inherits its peristence from the shock, our proposed solution builds on the general intuition from Stock and Watson (2018), showing that the inclusion of leads of the instrument can deliver valid inference under LP-IV.

3.5 Discussion

In the presence of persistence, a researcher has to determine first what object she wants to identify and, second, what estimation method to use. Table 2 summarizes the adjustments required in LPs and MA regressions depending on the choice of identification and estimation.

Regarding the decision on the estimation method, MA regressions and LPs have different strengths. LPs impose fewer restrictions and do not rely on a dimension-reduction approach. However, the fact that LPs impose fewer restrictions translates into less precise responses. On the contrary, the MA representation often imposes strong assumptions on the linearity of the underlying data generating process (the most typical case is the estimation of standard VARs). When this restriction on the functional form is true, then the dynamic response

Table 2: Adapting LPs and MA regressions when shocks are persistent

Object of interest / Method	LPs	MA regressions
Response as if no persistence ($\mathcal{R}(h)^*$)	include leads	no action needed
Response with persistence ($\mathcal{R}(h)$)	no action needed	replace x_t with ε_t

computed from the VAR through the MA representation results on more efficient and precise estimations. However, small mistakes in the functional form when specifying a VAR will be compounded throughout the response horizon, due to its iterative nature, contrary to what occurs in impulse-responses computed through LP.

Regarding the decision on the object that needs to be identified, the researcher has to take into account that standard MA and LPs methods identify different dynamic effects when shocks display persistence ($\gamma \neq 0$), as shown in the previous section. Would the researcher want to identify the response *as if* the shock were uncorrelated ($\mathcal{R}(h)^*$) or the response that includes the effect of persistence ($\mathcal{R}(h)$)? The identification of responses *as if* the shock were uncorrelated (implemented by LPs with leads or using MA) is perhaps the most conceptually appealing option for some applications.

First, a shock that generates responses as $\mathcal{R}(h)^*$, regardless of the persistence of the process, is akin to Ramey's definition of what an empirical shock should be. This is because persistence is an undesired feature of a (weakly-defined) shock, and the researcher may not want this to influence the estimation of dynamic responses.

Second, when the researcher wants to compare the effects of different shocks (e.g. whether fiscal or monetary policy is more effective in stimulating output, or how different are oil and technology shocks), these may have different underlying data generating processes. For example, it may be the case that fiscal shocks tend to show more persistence or that a given identification procedure tend to generate shocks with less persistence. To the extent that computing responses $\mathcal{R}(h)^*$ effectively standardizes the dynamic responses of shocks with different data generating processes, it may be desirable to employ LPs with leads or MA regressions.[26]

However, computing responses that contain the effect of persistence from the (weakly-defined) shock $\mathcal{R}(h)$ could still be informative in some contexts. A researcher interested in estimating the *most likely dynamic response* of a variable to a shock according to the historical data, may be interested in including all the data features of such shock (e.g. persistence). This argument is similar to the one posed by Fisher and Peters (2010) and Ramey and Zubairy (2018) to support the use of the cumulative multiplier (the ratio of the integral of the output response to that of the government spending response) to evaluate the effectiveness of fiscal

[26]This would be also the case when comparing the same shock, identified with the same methods, using data from different countries. Since the data generating process of the shock in each country may be different, the standardization provided by $\mathcal{R}(h)^*$ should become particularly useful to establish cross-country comparisons.

policies. If we consider the effects of a monetary policy (weakly-defined) shock that cuts the policy rate by 1 percentage point, it is important to note that if that shock displays persistence, then the total monetary policy action (the evolution of the nominal interest following the initial tightening) may be different to what would occur if the shock were iid.

Additionally, some researchers may not agree to the above definition of shocks. That is, they may define shocks as innovations that potentially can contain persistence.[27] In such applications, a researcher may consider the responses under standard LPs or using modified MA regressions, as shown in the last row of Table 2.

To sum up, we argue that the researcher may need to take decision on both the identification and the estimation of impulse-responses when shocks present persistence. The latter is determined by a bias-efficiency trade-off, while the former requires to consider what is the effect that the researcher is ultimately interested in. In the next section, we revisit some prominent empirical applications where the shocks are persistent and show that accounting for its persistence on the impulse responses may have a sizable impact.

4 Applications

In this section, we revisit some empirical papers that have used persistent shocks, and show how accounting for that persistence affects the estimated impulse responses.

4.1 Government spending shocks (Ramey and Zubairy (2018))

Ramey and Zubairy (2018), building on previous work by Ramey (2011) and Owyang et al. (2013), produce a series of announces about future defense spending between 1890q1-2014q1, scaled by previous quarter trend real GDP.[28] This series, plotted in panel D of Figure D2, has a positive autocorrelation of 18.4% (47.0% in the subsample after WWII).[29]

Ramey and Zubairy (2018) use LPs to estimate the response of output and government spending to a shock in future defense spending. We follow their same approach and sample and estimate the following equations for output (y_t) and government spending (g_t):

$$\begin{aligned} y_{t,h} &= \beta_h^y shock_t + \sum_{j=1}^{P} \rho_{j,h}^z z_{t-j} + \sum_{f=1}^{h} \gamma_{f,h} shock_{t+f} + \xi_t \\ g_{t,h} &= \beta_h^g shock_t + \sum_{j=1}^{P} \rho_{j,h}^z z_{t-j} + \sum_{f=1}^{h} \gamma_{f,h} shock_{t+f} + \varepsilon_t, \end{aligned} \tag{17}$$

[27] See for example Halac and Yared (2014).

[28] Ramey and Zubairy (2018) estimate trend GDP as sixth degree polynomial for the logarithm of GDP and multiplier by the GDP deflator. In fact, it is the use of the GDP deflator and trend GDP as a way to scale the shocks what seems to induce the persistence. The persistence is also present when the shock is scaled by previous-quarter GDP, as in Owyang et al. (2013).

[29] This positive autocorrelation is significant at a confidence level of 90% when considering standard errors that are robust to the presence of heteroskedasticity and persistence (with more than one lag) for the whole sample. For the subsample starting after WWII, the autocorrelation is significant at any level.

where z_t includes P lags of y_t, g_t and $shock_t$. Note that, following our method described in the previous section, we include h leads of the variable $shock_t$.[30] In particular, for each horizon h we include h leads.

To replicate Ramey and Zubairy (2018)'s estimates, we set $\gamma_{f,h} = 0$, $\forall\ f, h$. The black, solid line in Figure 4 represents the estimated responses of output (left panel) and government spending (right panel) to the shock.[31] The results closely resemble those in Ramey and Zubairy (2018) (Figure 5 of their paper).[32]

Next, we allow $\gamma_{f,h} \neq 0$. In the red lines in Figure 4, we observe that the responses change considerably when the leads are included. For example, after two years, output and government spending are 40% lower than in Ramey and Zubairy (2018)'s estimates.

Including or not leads also has implications for inference. The 95% confidence intervals when leads are included (shown in dashed lines in Figure 4) are substantially narrower than when they are not (grey areas in Figure D3). The latter are around 50% broader after two years, and more than twice as big after three years.

The dynamic responses of output and government spending are informative about the expected path of these variables after a shock. To obtain a measure of the efficiency of fiscal policy (i.e. the increase of output per each dollar increase in government spending), Ramey and Zubairy (2018) use the cumulative multiplier, computed as: [33]

$$M_{t,h} = \frac{\sum_{i=1}^{h} \beta_h^y}{\sum_{i=1}^{h} \beta_h^g}. \tag{18}$$

We find that this statistic is not substantially affected by persistence of the shock (Figure D4). Given that both output and government spending react similarly when including leads of the shock, taking the ratio of the two variables attenuates the differences between both specifications.[34]

Non-linear effects. We now investigate whether the effect of persistence in the shock can affect the responses in a non-linear setting, i.e. if government spending multipliers are

[30]We include lags of the shocks as in the original paper by Ramey and Zubairy (2018). Note that, as explained above, these lags do *not* account for the persistence of the shock, and hence their inclusion does not have a noticeable impact on the results.

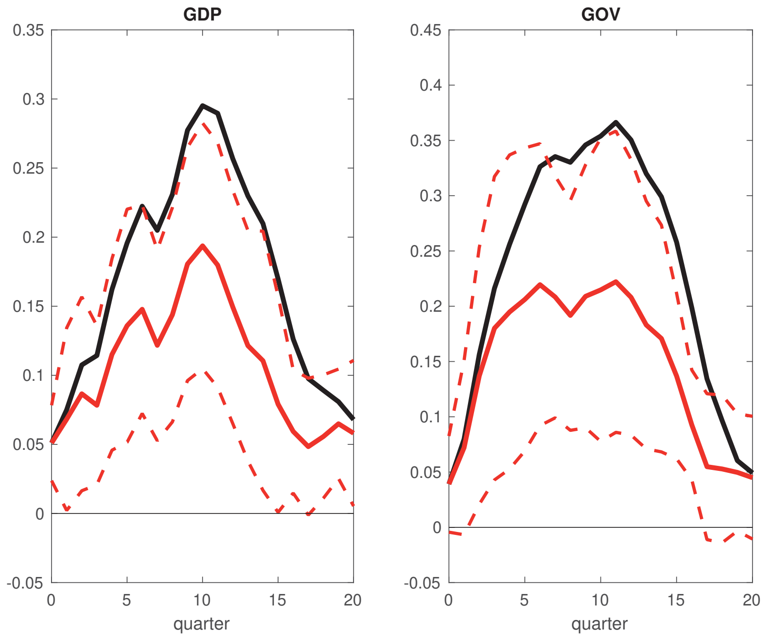
[31]Figure D3 also replicates the original 95% confidence intervals computed using the Newey-West correction.

[32]We drop the last h observations of the sample, so that the specifications with and without leads can be fully comparable. This does not have any discernible effect when replicating the original results from Ramey and Zubairy (2018).

[33]Ramey and Zubairy (2018) shows that the cumulative multiplier can be obtained in one step yielding identical results to those obtained combining equations (17) and (18).

[34]However, despite the effects of a serially-correlated and serially-uncorrelated shock are similar in terms of efficiency (multipliers), the fact that the expected responses are quantitatively different is very relevant from a policy -maker point of view. To see this point (which goes beyond considerations about normalizations of shocks), note that the the same shock could generate responses of output and government spending that differ in an an order of magnitude when accounting for the persistence, but still yield exactly the same multiplier. However, a higher response of government spending is likely to be relevant from a policy standpoint, as it can affect other important variables such as public debt or taxes.

Figure 4: Output and government spending responses, with and without leads



Black lines show the results of estimating the system (17) without including any lead (as in Ramey and Zubairy (2018)). Red solid lines represent the results of estimations when including h leads of the Ramey and Zubairy (2018) news variable (with 95% confidence intervals).

different in expansions and recessions.[35] For this, we follow Ramey and Zubairy (2018) and estimate a series of non-linear LPs:

$$\begin{aligned} x_{t+h} = S_{t-1} &\left[\alpha_{A,h} + \sum_{j=1}^{P} \rho_{A,j,h} z_{t-j} + \beta_{A,h} shock_t + \sum_{f=1}^{h} \delta_{A,f,h} shock_{t+f}\right] + \\ (1 - S_{t-1}) &\left[\alpha_{B,h} + \sum_{j=1}^{P} \rho_{A,j,h} z_{t-j} + \beta_{B,h} shock_t + \sum_{f=1}^{h} \delta_{B,f,h} shock_{t+f}\right] + \quad \xi_{t+h}, \end{aligned} \tag{19}$$

where x_t is either output or government spending and S_t is a binary variable indicating the state of the economy. When $S_t = 1$, the economy is booming and, when $S_t = 0$, the economy is in recession, which is defined as when the unemployment rate is above the threshold of 6.5. In this setting all the variables (including the constant), are allowed to have differential effects during expansions and recessions.

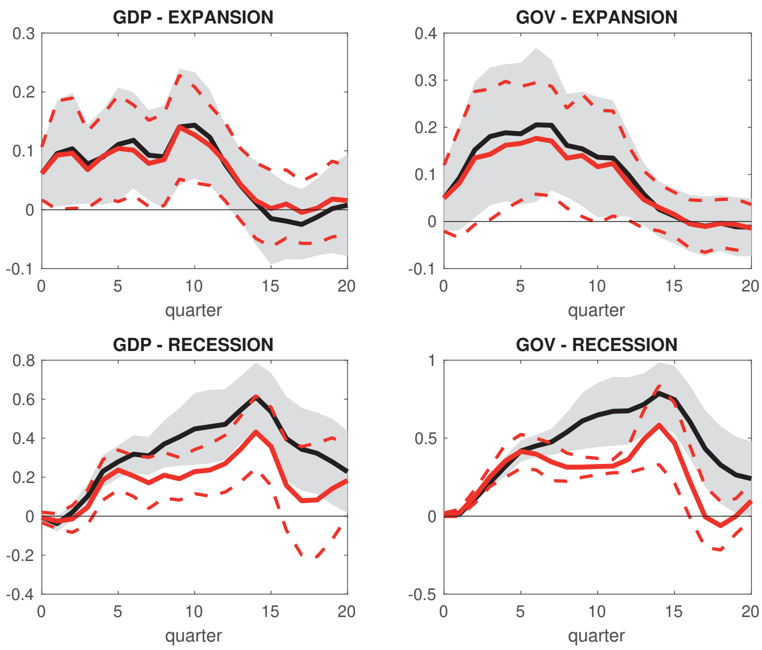
We first replicate the non-linear responses of output and government spending during booms and recessions obtained by Ramey and Zubairy (2018). Hence, we estimate equation (19) setting $\delta_{A,f,h} = \delta_{B,f,h} = 0 \ \forall f, h$. Our results, shown in Figure 5 in black lines, resemble very closely those from the authors. Next, we repeat the experiment accounting

[35]See Ramey (2019) for a recent summary of this debate. For example, an influential study by Auerbach and Gorodnichenko (2012) finds that government spending multipliers are higher during recessions using a non-linear VAR. Alloza (2018) highlights the role of the information used to define a period of recession, and finds that output responds negatively to government spending shocks in a post-WWII sample under different identification and estimation approaches.

for potential persistence, that is, including leads of the shock. The results are shown in red lines in Figure 5. While relatively similar in the case of expansions, the responses are quantitatively different during recessions. The estimates that include leads lie outside of the 95% confidence bands during much of the response horizon. The results suggest that ignoring the effect of persistence could yield responses during recessions that, after 2–3 years, are twice as large as the responses that account for the effect of persistence.

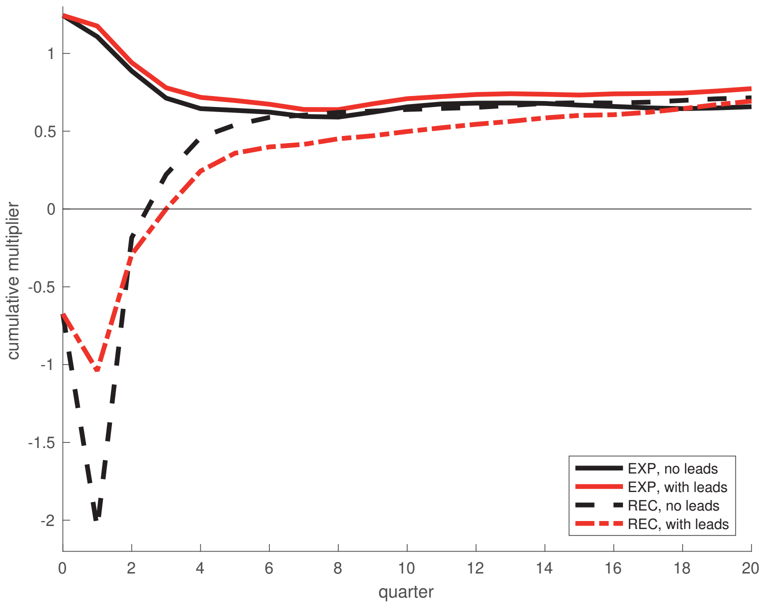
In Figure 6, we show how these responses map into estimates of non-linear fiscal multipliers. In the case of expansions, the results do not change much depending on whether the persistence is accounted for (red solid line) or not (black solid line). In either case, they resemble those in Ramey and Zubairy (2018) (see Figure 6 of their paper). In recessions, however, the results change substantially depending on whether the persistence is controlled for or not. If it is not (black solid line) the multiplier has a negative value upon impact and substantially falls in the following quarter to a value of -2. It becomes positive before the end of the first year, and fully converges to the value of the multiplier during expansions after six quarters. If the persistence is not controlled for (red dashed line), the multiplier shows a less puzzling behavior in the short run. After impact, the cumulative multiplier is -1 (instead of -2) and becomes positive after the first year. However, the multiplier during recessions remains lower than the multiplier during expansion for a much longer period. When the persistence is ignored this convergence is achieved after 6 quarters, as mentioned above. However, when including leads of the shock, this convergence is not fully reached during our

Figure 5: Responses during expansions and recessions, with and without leads



Black lines show the results from system of equations (19) without including any lead (as in Ramey and Zubairy (2018)). Grey areas represent 68 and 95% Newey-West confidence intervals for these estimates. Red solid lines represent the results of estimations when including h leads of the Ramey's news variable. Red dashed lines represent the 95% Newey-West confidence intervals for these estimates.

Figure 6: Government spending multiplier during expansions and recessions, with and without leads



The black solid and dashed lines show the cumulative multiplier during periods of expansion and recession, respectively, without including any lead (as in Ramey and Zubairy (2018)). The red solid and dashed lines show the cumulative multiplier during periods of expansion and recession, respectively, when including leads of the shock.

considered response horizon. These results suggest that during the short and medium-run the government spending multiplier could be lower during recessions than during expansions, and part of this difference may be attributable to the presence of persistence in the shock.

One of the main advantages of LPs is that they allow to accommodate non-linear settings, as those in equation (19). This is particularly useful since, contrary to threshold VARs, LPs do not impose any restriction on the evolution of state S_t (while non-linear VARs that interact the shock with a state dummy do assume that S_t remains fixed during the response horizon).The framework explained in the previous section allows to consider additional macroeconomic experiments that can help understand how restrictive this condition is. In particular, by including leads of the state S_t in equation (19) we are identifying the counterfactual response to a fiscal shock when the underlying state of the economy is not allowed to change (as in threshold VARs). We perform this experiment and report the multipliers during booms in recessions in green lines in Figure D5. We observe that when the state is not allowed to change, the multiplier during recessions is slightly higher in the short run, but essentially unchanged at medium and longer horizons. This exercise allows us to illustrate how the use of leads of variables in conjuction with LPs can help understand interesting counterfactual exercises.

4.2 Fiscal consolidations (Guajardo et al. (2014))

In this subsection we explore the relevance of our results in the context of episodes of fiscal consolidation, as produced in Guajardo et al. (2014). The authors employ a panel of OECD economies to analyze the response of economic activity to discretionary changes in fiscal policy motivated by a desire to reduce the budget deficit and not correlated with the short-term economic outlook.[36] As mentioned in Table 1, this measure of fiscal changes exhibits some degree of persistence.[37]

To explore the effects of persistence in this context, we compute the responses estimating a series of LPs:[38]

$$y_{i,t+h} = \mu_{h,i} + \lambda_{h,t} + \beta_{h,0} shock_{i,t} + \sum_{f=1}^{h} \beta_{h,f} shock_{i,t+f} + \boldsymbol{\beta}_{h,s} \boldsymbol{X}_{i,t} + \xi_{i,t+h}, \tag{20}$$

where $y_{i,t}$ is a measure of economic activity (either private consumption or real GDP). $\mu_{h,i}$ and $\lambda_{h,t}$ represent country and time fixed effects, respectively. X_t is a vector of variables that includes a lag of the shock, output and private consumption, and a deterministic trend. In our setting, responses to the fiscal shocks are given by the estimates of coefficients $\beta_{h,0}$ for different horizons h.

We first estimate equation (20) by setting $\beta_{h,f} = 0 \; \forall h, f$. The results, shown in black solid lines in Figure 7 qualitatively replicate the benchmark results of Guajardo et al. (2014), with a fiscal consolidation shock significantly reducing output during the first 6 years.[39]

Next, we estimate equation (20) but allow $\beta_{h,f} \neq 0$ (red lines in Figure 7). Three points are worth noting regarding these results. First, when accounting for the effects of persistence, the point estimates are smaller in absolute value. On average, the new responses are 35% lower during the first six years after the shock. Two years after a fiscal consolidation, output is almost 60% smaller when accounting for persistence (-0.2 *vs* -0.5).

Second, when including leads of the shock, the estimates are more precise, which translates into smaller confidence intervals (set at 90% as in the original paper of Guajardo et al. (2014)). During the first six years, these intervals are about 20% smaller on average in the specifications that include leads of the shock.

Third, these narrower intervals now include zero for most of the response horizon. Ignoring the persistence of the shock would lead to the conclusion that the output contraction after

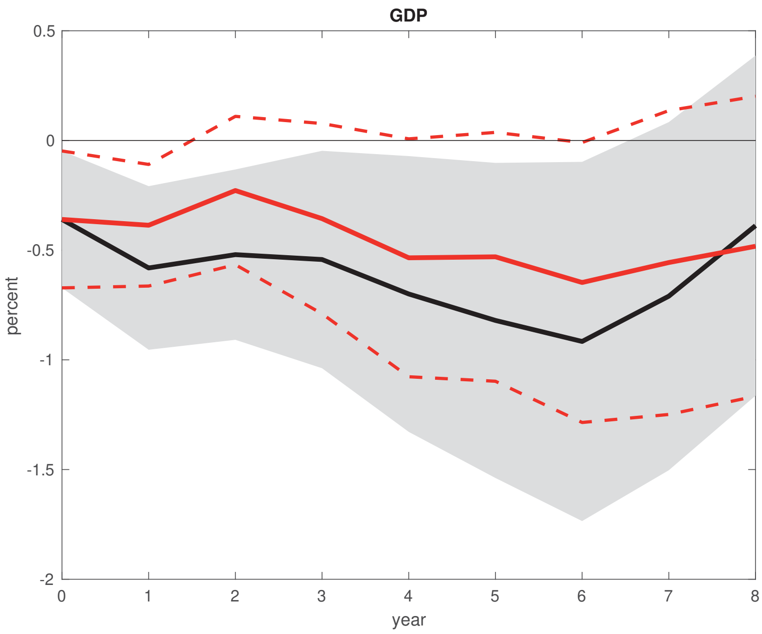
[36] A detailed description of these shocks can be found in Devries et al. (2011).

[37] Regressions of the fiscal consolidations measure (expressed as % of GDP) on its own lags and including time and country fixed effects reveal persistence in the previous two or three years (depending on the number of lags included). Intuitively, some degree of persistence is expected in these series since they often involved multi-year plans, as noted in Alesina et al. (2015) and Alesina et al. (2017).

[38] Note that Guajardo et al. (2014) do not construct responses using LPs and hence their computed responses do not show the effect of persistence, as noted in the previous section. There are, however, a number of studies that employ their fiscal consolidations dataset with LPs (see, for example, Barnichon and Matthes (2017) or Goujard (2017)).

[39] Guajardo et al. (2014) focus on the dynamic effects of output and private consumption during 6 years after the shock. We also compute results for private consumption, shown in Figure D6 in Appendix D. As in the original paper, we also find a significant reduction in this variable during the first 6 years after a consolidation shock.

Figure 7: Output response to a fiscal consolidation shock, with and without leads



Black lines show the results from equation (20) with output as dependent variable and setting $\beta_{h,f} = 0 \; \forall h, f$, i.e. without including any leads of the shock. Grey areas represent 90% Newey-West confidence intervals for these estimates (same interval as reported in Guajardo et al. (2014)). Red solid lines represent the results of estimations when allowing $\beta_{h,f} \neq 0$ and including h leads of the consolidations variable.

a fiscal consolidation is significant throughout the six years after the shock. However, when accounting for persistence, the effect of the shock is significant only during the first year after the shock, while it seems less plausible to conclude that the effect is statistically different from zero during the rest of the response horizon.

4.3 Tax shocks (Romer and Romer (2010))

What happens when including leads of non-persistent shocks? In this section we conduct a *placebo* test based on Romer and Romer (2010), who investigate the output effects of legislated tax changes. Romer and Romer (2010) identify exogenous changes in tax revenues by classifying fiscal reforms according to their motivation (i.e. whether or not they are the response to changing macroeconomic conditions). As discussed in Section 2, it is the only shock considered here for which we unambiguously fail to reject the null hypothesis of no persistence. Hence, the inclusion of leads of the shock should not have a discernible impact on the estimation of dynamic responses.

To empirically demonstrate this, we estimate the response of output to exogenous tax changes following Romer and Romer (2010). We adapt the original estimation from the authors (a single equation with no controls estimated by OLS) to the LPs setting:[40]

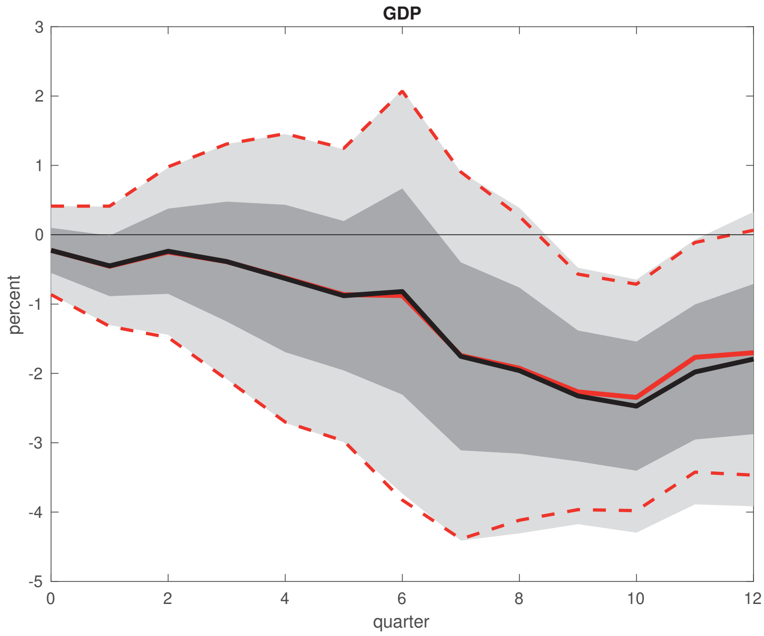
$$\frac{y_{t+h} - y_{t-1}}{y_{t-1}} = \beta_{h,0} shock_t + \sum_{f=1}^{h} \beta_{h,f} shock_{t+f} + \xi_{t+h}. \tag{21}$$

[40] Adding controls such as lags of output or the own shock do not affect the obtained results shown next.

In our first exercise, we set $\beta_{h,f} = 0\ \forall h, f$ in equation (21) to replicate the results from Romer and Romer (2010). The results are shown Figure 8 (black lines). The response of output is similar to that in Romer and Romer (2010): it falls persistently after a tax hike of 1% of GDP, with a peak effect reached in the 10th quarter.[41]

Next, we allow for $\beta_{h,f} \neq 0$. The results, shown in Figure 8 (red lines), suggest that the inclusion of leads does not significantly affect the results. The point estimations with and without leads of the shock overlap each other for most of the response horizon and only diverge slightly during the quarters 8 to 11th.

Figure 8: Response of output to Romer and Romer (2010) tax shocks, with and without leads



Black solid line shows the responses to a tax shock estimated from equation(21) with $\beta_{h,f} = 0$, i.e. without including any lead. Grey areas represent 68 and 95% Newey-West confidence intervals for these estimates. Red solid line shows the responses to a tax shock estimated from equation (21) with $\beta_{h,f} \neq 0$ and including h leads of the shock. Red dashed lines represent 95% Newey-West confidence intervals for these estimates.

While, given the results of Table 1 we should not expect a change in the point estimates (which we have corroborated) the same cannot be say about issues regarding inference. However, Figure 8 shows that confidence bands are not distinguishable between both specifications during the first seven quarters and differ only slightly afterwards.

In sum, this placebo exercise is reassuring in that the inclusion of leads only matters when the explanatory variable displays some persistence. These results suggest that including leads in LPs is a conservative way to address the effects of persistence when there is a suspicion that the shock is persistent.

[41]The difference with the original estimations from Romer and Romer (2010) are only quantitative: the peak tax multiplier is about 3 in the 10th quarter. Our estimations suggest a peak multiplier of 2.25 also reached in the same quarter.

5 Conclusions

In this paper, we have shown that persistence results in the estimation of different responses when using LPs *versus* traditional methods based on MA regressions. In particular, for a researcher interested in the response *as if* the shock were not persistent, MA regressions yield the desired object, but LPs need to be adapted. We have proposed a method based on the inclusion of leads of the shock to address this issue.

Accounting for persistence is particularly relevant from a policy standpoint. The comparison of the effectiveness of different shocks should be based on the assumption that the data generating processes underlying the shocks are comparable. Otherwise, the researcher could inadvertently overestimate the effect of shocks that shows with high persistence. However, there might be applications where including all the features of the data (e.g. persistence) may be advantageous. The aim of our paper is to show that these features are treated differently depending on the method used to identify the dynamic effects.

The use of leads has an additional advantage in LPs beyond addressing concerns about persistence. As it has been shown, the inclusion of leads of a variable effectively shuts down the transmission channel that operates through that variable due to its persistence. For example, one may be interested in the effects of monetary policy shocks on output due to a particular instrument while holding other variables constant (e.g. changes to fiscal policy). In the context of LPs, leads of a selected variable (e.g. tax changes) will deliver responses holding that variable constant. This methodology allows to separate the direct (due to the impact through the regressor of interest) and indirect effects (due to other variables in the regression). This has often been used in the context of VARs, by imposing restrictions on the coefficients of selected impulse responses.[42] The inclusion of leads achieves a similar goal in LPs, hence allowing to construct interesting macroeconomic experiments. We leave these questions for future research.

[42] See for example: Bernanke et al. (1997), Sims and Zha (2006), or Bachmann and Sims (2012).

References

Alesina, A., Barbiero, O., Favero, C., Giavazzi, F., and Paradisi, M. (2017). The effects of fiscal consolidations: Theory and evidence. Working Paper 23385, National Bureau of Economic Research.

Alesina, A., Favero, C., and Giavazzi, F. (2015). The output effect of fiscal consolidation plans. *Journal of International Economics*, 96:S19–S42.

Alloza, M. (2018). Is Fiscal Policy More Effective in Uncertain Times or During Recessions? Working Paper 1730, Banco de España.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58(2):277–297.

Arezki, R., Ramey, V. A., and Sheng, L. (2017). News shocks in open economies: Evidence from giant oil discoveries. *Quarterly Journal of Economics*, 132(1):103–155.

Auerbach, A. and Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Bachmann, R. and Sims, E. R. (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics*, 59(3):235 – 249.

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

Barnichon, R. and Brownlees, C. T. (2018). Impulse response estimation by smooth local projections. *Review of Economics and Statistics*, forthcoming.

Barnichon, R. and Matthes, C. (2017). Understanding the Size of the Government Spending Multiplier: It's in the Sign. Working paper, Mimeo.

Bernanke, B. S., Gertler, M., Watson, M., Sims, C. A., and Friedman, B. M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity*, 1997(1):91–157.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3):623–685.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Choi, I. and Kurozumi, E. (2012). Model selection criteria for the leads-and-lags cointegrating regression. *Journal of Econometrics*, 169(2):224–238.

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

Cloyne, J. (2013). Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom. *American Economic Review*, 103(4):1507–28.

Cloyne, J. and Hürtgen, P. (2016). The Macroeconomic Effects of Monetary Policy: A New Measure for the United Kingdom. *American Economic Journal: Macroeconomics*, 8(4):75–102.

Coibion, O. (2012). Are the Effects of Monetary Policy Shocks Big or Small? *American Economic Journal: Macroeconomics*, 4(2):1–32.

Devries, P., Guajardo, J., Leigh, D., and Pescatori (2011). A New Action-Based Dataset of Fiscal Consolidation. Working paper, IMF Working Paper No. 11/128.

Faust, J. and Wright, J. H. (2011). Efficient prediction of excess returns. *Review of Economics and Statistics*, 93(2):647–659.

Fieldhouse, A., Mertens, K., and Ravn, M. O. (2017). The macroeconomic effects of government asset purchases: evidence from postwar US housing credit policy. *The Quarterly Journal of Economics*.

Fisher, J. D. and Peters, R. (2010). Using Stock Returns to Identify Government Spending Shocks. *The Economic Journal*, 120(544):414–436.

Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Geweke, J. F. and Singleton, K. J. (1981). Maximum likelihood" confirmatory" factor analysis of economic time series. *International Economic Review*, pages 37–54.

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

Goujard, A. (2017). Cross-Country Spillovers from Fiscal Consolidations. *Fiscal Studies*, 38(2):219–267.

Guajardo, J., Leigh, D., and Pescatori, A. (2014). Expansionary austerity? International evidence. *Journal of the European Economic Association*, 12(4):949–968.

Halac, M. and Yared, P. (2014). Fiscal rules and discretion under persistent shocks. *Econometrica*, 82(5):1557–1614.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

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

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Mertens, K. and Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks. *American Economic Journal: Economic Policy*, 4(2):145–81.

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.

Miranda-Agrippino, S. and Ricco, G. (2018). The transmission of monetary policy shocks. *CEPR Discussion Paper No. DP13396*.

Owyang, M. T., Ramey, V. A., and Zubairy, S. (2013). Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data. *American Economic Review*, 103(3):129–34.

Plagborg-Møller, M. and Wolf, C. K. (2019). Local Projections and VARs Estimate the Same Impulse Responses. *Manuscript, Princeton University*.

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

Ramey, V. and Shapiro, M. (1998). Costly Capital Reallocation and the Effects of Government Spending. In *Carnegie-Rochester Conference Series on Public Policy*, volume 48, pages 145–194. Elsevier.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier.

Ramey, V. A. (2019). Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research? *Journal of Economic Perspectives*, 33(2):89–114.

Ramey, V. A. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from US historical data. *Journal of Political Economy*, 126(2):850–901.

Romer, C. D. and Romer, D. H. (1989). Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. *NBER Macroeconomics Annual*, 4:121–170.

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.

Sims, C. A. and Zha, T. (2006). Does monetary policy generate recessions? *Macroeconomic Dynamics*, 10(2):231–272.

Stock, J. and Watson, M. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica*, 61(4):783–820.

Stock, J. H. and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal*, 128(610):917–948.

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.

Teulings, C. N. and Zubanov, N. (2014). Is economic recovery a myth? Robust estimation of impulse responses. *Journal of Applied Econometrics*, 29(3):497–514.

Online Appendices

A Proofs

A.1 Proof of Proposition 1

Consider equation (5) (rewritten here for convenience):

$$y_{t+h} = \delta_h x_t + \xi_{t+h}, \tag{A.1}$$

where $\delta_h = \mathcal{R}(h)^{LP}$ represents the impact of variable x_t on y_{t+h} (the response function). Since δ_h is the linear projection coefficient of equation (A.1):

$$\delta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)}. \tag{A.2}$$

The dynamic effect of x_t on y_{t+h} can also be obtained from MA regressions as in equation (4):

$$y_t = \theta_0 x_t + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \theta_3 x_{t-3} + \theta_4 x_{t-4} \ldots + u_t.$$

Since this expression holds $\forall t$, it can be written as:

$$y_{t+h} = \theta_0 x_{t+h} + \theta_1 x_{t+h-1} + \theta_2 x_{t+h-2} + \theta_3 x_{t+h-3} + \ldots + \theta_h x_t + u_t,$$

where the coefficient $\theta_h = \mathcal{R}(h)^{MA}$ represents the impulse response in period h, obtained from:

$$\theta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\ldots,x_{t+h}}. \tag{A.3}$$

When $\gamma = 0$ in the process described by system (1), we have that $x_t = \varepsilon_t \sim white\ noise(\mu_\varepsilon,\ \sigma^2_\varepsilon)$, so:

$$\theta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\ldots,x_{t+h}} = \frac{cov(y_{t+h}, x_t)}{var(x_t)}.$$

In this case, $\delta_h = \theta_h$ and LPs and MA yield the same responses: $\mathcal{R}(h)^{LP} = \mathcal{R}(h)^{MA}\ \forall h$. Note that $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^*\ \forall h, \gamma$ since (under linearity):

$$\mathcal{R}(h)^* = \frac{\partial y_{t+h}}{\partial x_t}\Big|_{x_{t+1},\ldots,x_{t+h}} = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\ldots,x_{t+h}}.$$

When $\gamma \neq 0$, equation (A.2) becomes (using the equations in system (1)):

$$\delta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)} = \frac{cov(\delta x_{t+h} + u_{t+h}, x_t)}{var(x_t)} = \delta \frac{cov(x_{t+h}, x_t)}{var(x_t)} = \delta\gamma^h, \tag{A.4}$$

using the expression:

$$x_{t+h} = \gamma^h x_t + \sum_{j=0}^{h-1} \gamma^j \varepsilon_{t+h-j}. \tag{A.5}$$

However, the dynamic response obtained from MA regressions is:

$$\theta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\dots,x_{t+h}} = \delta\frac{cov(x_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\dots,x_{t+h}}.$$

When $h = 0$, the above expression becomes $\theta_0 = \delta\frac{cov(x_t,x_t)}{var(x_t)} = \delta$. For $h > 0$, we have that $\theta_h = \delta\frac{cov(x_{t+h},x_t)}{var(x_t)}\Big|_{x_{t+1},\dots,x_{t+h}} = 0$. This shows that when $\gamma \neq 0$, we have that $\mathcal{R}(h)^{LP} = \mathcal{R}(h)^{MA}$ if and only if $h = 0$. ■

A.2 Proof of Proposition 2

Consider equation (9) (rewritten here for convenience):

$$y_{t+h} = \delta_{h,0}x_t + \delta_{h,1}x_{t+1} + \xi_{t+h}, \tag{A.6}$$

where $\delta_{h,0} = \mathcal{R}(h)^F$ represents the impact of the shock x_t on y_{t+h} when including leads of the former. Since $\delta_{h,0}$ is the linear projection coefficient of equation (A.6), then:

$$\delta_{h,0} = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1}} = \delta\frac{cov(x_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1}}. \tag{A.7}$$

Note that the data generating process in system (1) considers that x_t is an AR(1) so it can be represented in terms of x_{t+1} (see equation (A.5)).[1] Then, we have that MA regressions and LPs with leads recover the same object:

$$\theta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\dots,x_{t+h}} = \delta\frac{cov(x_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1}} = \delta_{h,0}.$$

To see this, note that in period $h = 0$ we have that:

$$\theta_0 = \frac{cov(y_t, x_t)}{var(x_t)}\Big|_{x_{t+1},\dots,x_{t+h}} = \frac{cov(y_t, x_t)}{var(x_t)} = \delta = \delta_{0,0}.$$

In periods $h > 0$, we can rewrite equation (A.7) as:

$$\delta_{h,0} = \delta\frac{cov(x_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1}} = \delta\gamma^{h-1}\frac{cov(x_{t+1}, x_t)}{var(x_t)}\Big|_{x_{t+1}} = 0. \tag{A.8}$$

Similarly, equation (A.3) becomes:

$$\theta_h = \frac{cov(y_{t+h}, x_t)}{var(x_t)}\Big|_{x_{t+1},\dots,x_{t+h}} = \delta\gamma^{h-1}\frac{cov(x_{t+1}, x_t)}{var(x_t)}\Big|_{x_{t+1}} = 0. \tag{A.9}$$

So we have $\mathcal{R}(h)^F = \mathcal{R}(h)^{MA}\ \forall\ h,\gamma$. And we know that $\mathcal{R}(h)^{MA} = \mathcal{R}(h)^*$, from the section above. ■

[1] Note also that the results easily generalize to cases when x_t is an autoregressive process of higher order.

A.3 Proof of Proposition 3

Consider a version of equation (11) rewritten here for convenience:

$$y_t = \tilde{\theta}_0 \varepsilon_t + \tilde{\theta}_1 \varepsilon_{t-1} + \tilde{\theta}_2 \varepsilon_{t-2} + \tilde{\theta}_3 \varepsilon_{t-3} + \tilde{\theta}_4 \varepsilon_{t-4} \ldots + u_t, \tag{A.10}$$

where $\tilde{\theta}_0 = \mathcal{R}(h)^{MA-per}$ represents the impact of variable ε_t on y_{t+h}. Note that ε_t is not observable but can be obtained if we know the data generating process described in system (1). Since equation (A.10) represents the linear projection of y_t on ε_t and its lags, with $\varepsilon_t \sim iid(\mu_\varepsilon, \sigma_\varepsilon^2)$, we have:

$$\tilde{\theta}_h = \frac{cov(y_{t+h}, \varepsilon_t)}{var(\varepsilon_t)}\bigg|_{\varepsilon_{t+1},\ldots,\varepsilon_{t+h}} = \frac{cov(y_{t+h}, \varepsilon_t)}{var(\varepsilon_t)} = \delta \frac{cov(x_{t+h}, \varepsilon_t)}{var(\varepsilon_t)}. \tag{A.11}$$

This expression is equivalent to equation (A.2) which implies that $\tilde{\theta}_h = \delta_h$ and $\mathcal{R}(h)^{MA-per} = \mathcal{R}(h)^{LP}$. To see this, substitute for x_{t+h} in equation (A.11) using expression (A.5) and system (1):

$$\tilde{\theta}_h = \delta\gamma^h \frac{cov(x_t + \sum_{j=0}^{h-1} \gamma^j \varepsilon_{t+h-j}, \varepsilon_t)}{var(\varepsilon_t)} = \delta\gamma^h \frac{cov(x_t, \varepsilon_t)}{var(\varepsilon_t)} = \delta\gamma^h. \tag{A.12}$$

Note that the above expression yields the same result as equation (A.4), which shows that $\tilde{\theta}_h = \delta_h \ \forall \ h$. ■

B Further results

B.1 Responses in LPs using variables adjusted for serial correlation

An apparent potential alternative to the use of leads proposed in the main text might be to adjust the shock x_t so that it does not display persistence (e.g. by regressing x_t on its own lags and using the resulting residual). Once the persistence is removed, one may expect the dynamic responses not to include the effect due to the persistence of the shock. However, this is not the case in a LPs setting, as we show next.

Consider the case where we obtain a variable adjusted for serial correlation: $\varepsilon_t = x_t - \gamma x_{t-1}$, as shown in equation (8). Then, ε_t can be used as substitute of the original shock x_t. Assuming $B_1 = 0$ in system (12) (for simplicity) consider the following series of LPs:

$$y_{t+h} = \rho_h y_{t-1} + \lambda_h \varepsilon_t + \xi_{t+h}. \tag{B.1}$$

To obtain the dynamic responses of y_t to the shock ε_t (adjusted for persistence), we rewrite the first equation in system (12) as a function of ε_t and compute the relevant partial derivatives. For the cases of $h = 0$ and $h = 1$ these are:

$$\begin{aligned} \lambda_0 = \frac{\partial y_{t+1}}{\partial \varepsilon_t} &= B_0 \\ \lambda_1 = \frac{\partial y_{t+1}}{\partial \varepsilon_t} &= \rho \frac{\partial y_t}{\partial \varepsilon_t} + B_0 \frac{\partial x_{t+1}}{\partial \varepsilon_t} = \rho B_0 + B_0 \gamma = B_0(\gamma + \rho). \end{aligned} \tag{B.2}$$

That is, even after correcting for the persistence in shock x_t, conventional LPs yield responses that still contain the effect of persistence of the shock.

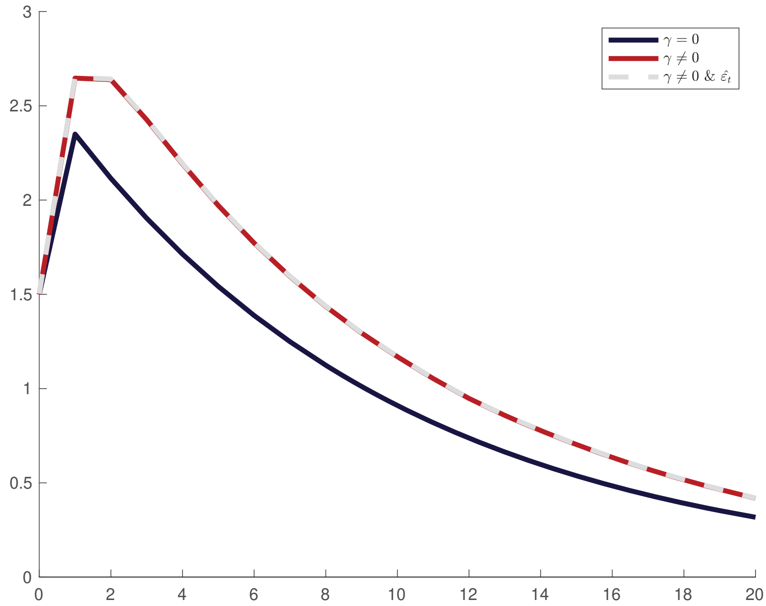
While this result may seem counter-intuitive, it arises from the fact that LPs do not have an explicit dynamic structure as an MA representation. Hence, removing the persistence from x_t does not eliminate its effect on y_{t+1}, y_{t+2}, etc.

To empirically show this point, we simulate series of y_t and x_t following system (12) and the calibration used in Section 3.3 (we now allow $B_1 \neq 0$). We then obtain the residuals $\hat{\varepsilon}_t$ as an estimate of ε_t described above and estimate the following equation:

$$y_{t+h} = \rho y_{t-1} + \lambda_{h,0} \hat{\varepsilon}_t + \lambda_{h,1} \hat{\varepsilon}_{t-1} + \xi_{t+h}. \tag{B.3}$$

Results are shown in Figure B1. The simulations corroborate the above results and we find that the use of a variable adjusted for serial correlation as $\hat{\varepsilon}_t$ in equation (B.1) fails to retrieve an impulse response as the one obtained when $\gamma = 0$ in equation (12).

Figure B1: Simulated responses using $\hat{\varepsilon}_t$



This figure shows the response of a simulated outcome variable to a shock with different degrees of persistence. The dark blue line shows the results of estimating equation (B.3) assuming $\gamma = 0$ in equation (12). The red line shows the same estimation when $\gamma = 0.2$. The dashed grey line shows the response when including a predicted regressor where persistence has been removed as explanatory variable (as in equation (B.1)).

B.2 Including the shocks as endogenous variables in a VAR

A researcher may consider including a shock with persistence as an endogenous variable in a VAR. Does this approach eliminate the effect of the persistence of the shock on the impulse responses? A VAR, since it explicitly models the persistence of the shock, includes this effect in the estimated impulse responses and, hence, yields the same dynamic effects as LPs (contrary to what is obtained when including the shock as a MA structure within a VAR).[1]

To see this in an intuitive way, consider the data generating process given by system (12) and rewritten here for convenience (with a slightly different notation):

$$\begin{aligned} y_t &= \rho y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + \varepsilon_t^y \\ x_t &= \gamma x_{t-1} + \varepsilon_t^x. \end{aligned} \tag{B.4}$$

This process can be recast as a structural VAR of the form $\boldsymbol{A}_0 \boldsymbol{Y}_t = \boldsymbol{B}^* \boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t$, with:

$$\begin{bmatrix} 1 & 0 \\ -\delta_0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \gamma & 0 \\ \delta_1 & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix}. \tag{B.5}$$

[1]Romer and Romer (2010), Ramey (2011), and Bloom (2009) are examples of studies that include shocks as endogenous variables in a VAR. Plagborg-Møller and Wolf (2019) formally show that VARs and LPs identify the same impulse responses. Here we illustrate that when one of the endogenous variables in the VAR is a persistent shock, this effect will be carried over to the dynamic responses.

An econometrician would estimate the following reduced-form VAR:

$$\boldsymbol{Y}_t = \boldsymbol{B}\boldsymbol{Y}_{t-1} + \boldsymbol{u}_t, \tag{B.6}$$

where $\boldsymbol{B} = \boldsymbol{A}_0^{-1}\boldsymbol{B}^*$; and $\boldsymbol{u}_t = \boldsymbol{A}_0^{-1}\boldsymbol{\varepsilon}_t$ are reduced-form residuals. Since the data generating process given by equation (B.4) already incorporates restrictions on the contemporaneous behavior of the variables, a researcher may identify the structural impulse responses by computing the Choleski decomposition (when x_t is ordered first) of the variance-covariance matrix of reduced-form residuals $\boldsymbol{u}_t$.

However, note that, when $\gamma \neq 0$, the response of y_t to ε_t^x will include an effect due to the persistence of the shock x_t. Intuitively, consider the case of $\rho = \delta_1 = 0$. In this scenario, a researcher may be interesting in recovering a *one-off* shock to x_t. However, the response of y_t will be given by $\mathcal{R}(h)^{VAR} = \delta_0 \sum_{k=0}^{\infty} \gamma^k \varepsilon_{t-k}^x$, that is, the *one-off* shock will still have effects along the response horizon because of the peristence of x_t (when $\gamma \neq 0$).

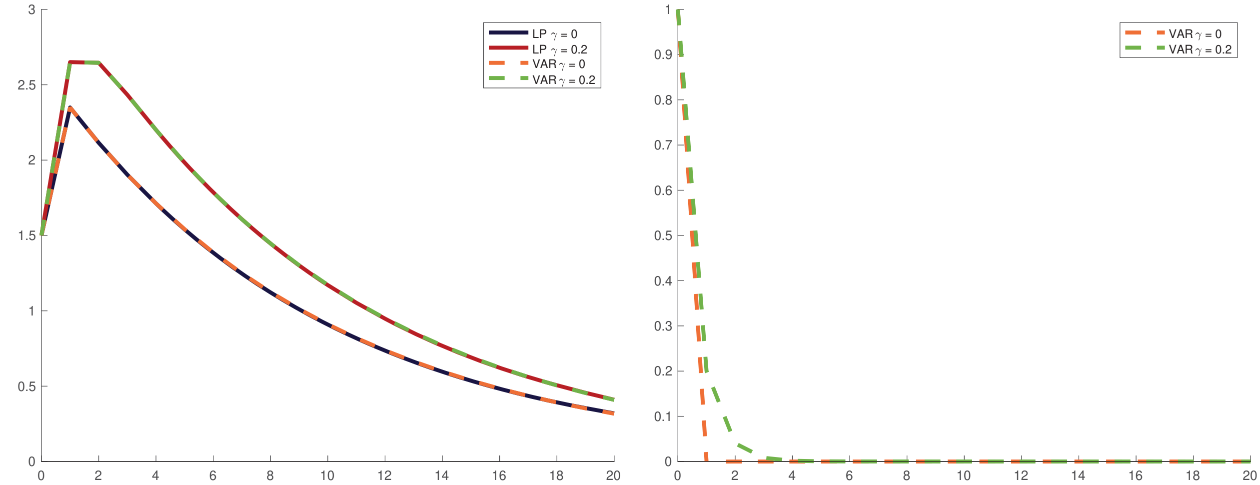
To see this point in a more general way, consider the same calibration as used in the main text (i.e. $\rho = 0.9$, $\delta_0 = 1.5$, $\delta_1 = 1$, and either $\gamma = 0$ or $\gamma = 0.2$). We compute the impulse responses of variables y_t and x_t (the measured shock) to ε_t^x, shown in Figure B2. We also estimate the same impulse-response functions for y_t using LPs, as in Section 3.3.

The results illustrate that, regardless of the value γ, a VAR that considers x_t as endogenous variable and LPs estimate the same impulse responses. When there is some persistence in the shock x_t, both the VAR (that considers x_t as an endogenous variable) and LPs include a dynamic effect due to the persistence of x_t.

Importantly, when x_t displays persistence, the response function estimated by a VAR will vary depending on whether x_t is included as an endogenous variable (as shown before) or an exogenous one (e.g. with a MA structure). To show this point, Figure B3 displays the response of y_t when (i) x_t is included as an endogenous variable in the VAR, or (ii) when estimating a regression of y_t on x_t and a lag of both variables. As it can be seen, considering x_t as an exogenous regressor always delivers the same dynamic responses as if x_t were serially uncorrelated, regardless of the actual value of γ.

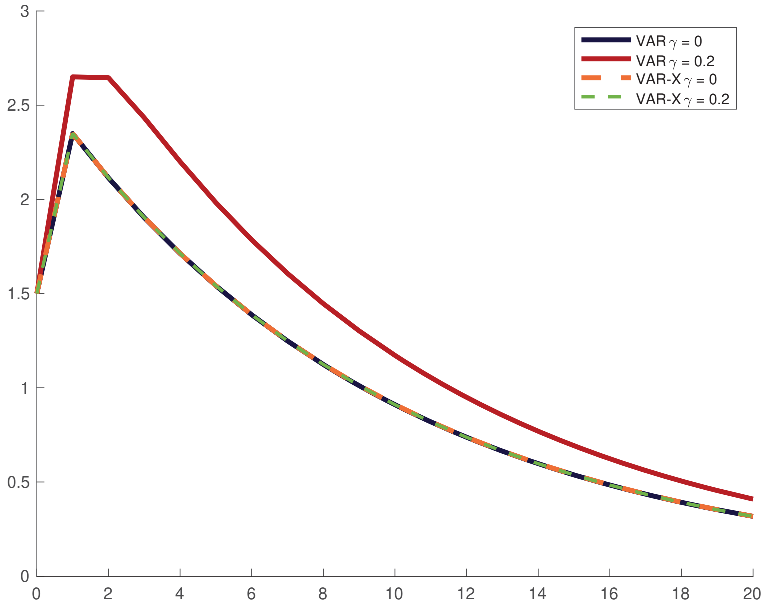
Figure B2: Responses in a VAR to ε_t^x

Panel A) Response of y_t to ε_t^x Panel B) Response of x_t to ε_t^x



The figure shows the VAR responses of y_t (Panel A) and x_t (Panel B) to ε_t^x estimated from (B.6), under different assumptions of the persistence parameter γ: dashed orange lines for responses when there is no persistence ($\gamma = 0$) and dashed green lines for responses when there is persistence ($\gamma = 0.2$). For reference, Panel A also displays responses from the same DGP estimated using LPs for the cases of $\gamma = 0$ (solid blue line) and $\gamma = 0.2$ (solid red line).

Figure B3: Responses when the shock is included as an endogenous or exogenous variable in a VAR



The figure shows the VAR responses of y_t to ε_t^x estimated from (B.6) under different assumptions of the persistence parameter γ and two different specifications: solid lines display the responses when the shock is included as an endogenous variable in the VAR and dashed line shows the responses when the same shock is included as an exogenous variable in the VAR.

B.3 Alternative simulation: using the persistence from an actual shock

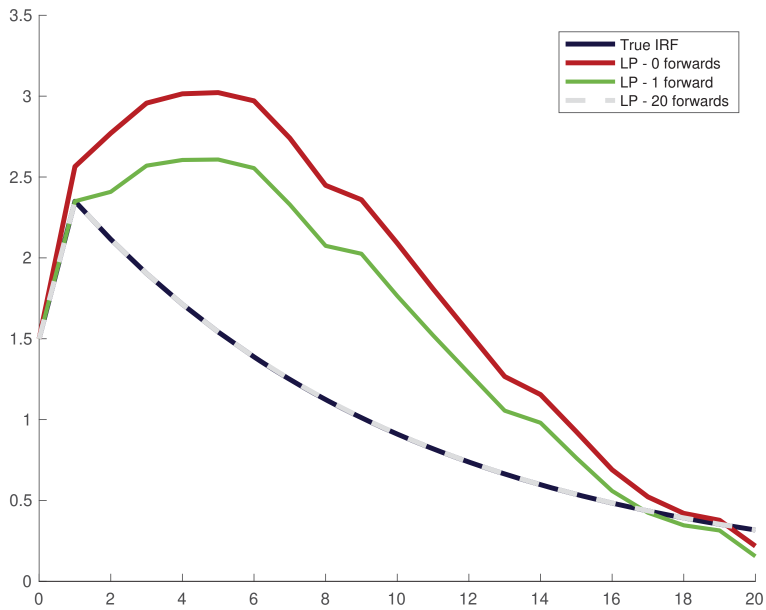
In this subsection we compute the impulse response of a simulated variable y_t to a shock x_t with the following DGP:

$$y_t = \rho y_{t-1} + B_0 x_t + B_1 x_{t-1} + u_t, \tag{B.7}$$

where x_t is the actual government spending shock from Ramey and Zubairy (2018) as shown in Panel D of Figure D2. u_t is a random variable following $u_t \sim \mathcal{N}(0, 1)$. We set $\rho = 0.9$, $B_0 = 1.5$, and $B_0 = 1$.

Equation (B.7) is simulated for 497 periods (the length of Ramey and Zubairy (2018)'s shock), and we then compute the relevant IRFs. We repeat this process 10,000 times, and compute the average impulse responses across all repetitions. The results are shown in Figure B4.

Figure B4: Simulated responses using LPs with persistence from an actual shock



This figure shows the response of a simulated outcome variable to the government spending shock from Ramey and Zubairy (2018). The dark blue line is the theoretical impulse-response to a shock that shows no persistence. The red line shows the LPs estimation of the impulse-response to the Ramey and Zubairy (2018) without including any lead. Green line repeats the same estimation adding one lead. Dashed grey line shows the response when including 20 leads.

When computing the dynamic response with standard LPs (i.e. without including any lead), the estimates diverge from the expected response when the shock has no persistence (distance between read and dark blue lines in Figure B4). Adding one lead improves the estimates, bringing the impulse-response into line with the theoretical response in the first period (green line). The accuracy of the impulse-response converges to the theoretical response when more leads are included. When we include as many leads as periods in the response horizon (20), the dynamic response estimated from LPs using the actual shock (with persistence) is equivalent to the response to a non-serially correlated shock (dashed grey line).

C Additional empirical applications: The effects of monetary policy

In this appendix we explore how the persistence found in some measures of monetary policy shocks can affect the dynamic responses estimates using LPs.

C.1 Romer and Romer (2004)

We first consider the measure of monetary policy shocks produced by Romer and Romer (2004). The authors identify exogenous monetary policy changes following a three-step procedure. First, they follow narrative methods to identify the Federal Reserve's intentions for the federal funds rate around FOMC meetings. Second, they regress the resulting measure on the Federal Reserve's internal forecasts (Greenbook) to account for all relevant information used by the Fed. Lastly, the series is aggregated from FOMC frequency to monthly frequency.

As shown in Table 1, the resulting measure displays some degree of persistence.[1] Interestingly, the correlogram of the series seems to show a pattern consistent with *negative* (see Panel E in Figure D1). This implies that standard LPs that do not account for in the shock should underestimate the dynamic effects.

Romer and Romer (2004) estimate the response of output to the monetary policy shock using a lag-distributed regression of the log of industrial output and the measure of monetary policy shocks. Here, we adapt the estimation to a LPs setting following the exact data and specification from Ramey (2016) (adapted in turn from Coibion (2012)) for the original sample of 1969m3-1996m12):

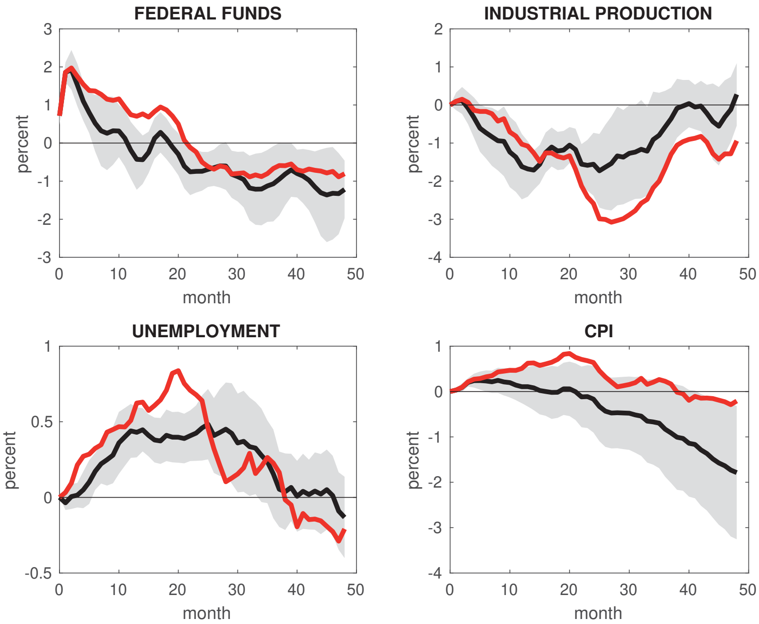
$$y_{t+h} = \beta_{h,0} shock_t + \boldsymbol{\theta_h(L) x_t} + \beta_{h,f} \sum_{f=1}^{h} shock_{t+f} + \xi_{t+h}, \tag{C.1}$$

where y_t is either the federal funds rates, the log of industrial production, the log of consumer price index, the unemployment rate, or the log of a commodity price index, and $shock_t$ is the original Romer and Romer (2004) measure of monetary policy shocks. The regressions include a set of controls $\boldsymbol{x}_t$, with two lags and the contemporaneous values of all dependent variables, and two lags of the shock. By including the contemporaneous values of the the dependent variables, we are implementing the recursiveness assumption often used in VARs to identify monetary policy shocks.[2]

[1] The degree of persistence is higher when using the updated series produced by Coibion (2012).

[2] This assumption implies that the monetary shock does not affect macroeconomic variables (such as output, prices, employment...) contemporaneously, and monetary variables (e.g. money stock, reserves...) do not affect the federal funds rates within a month. See Christiano et al. (1999) for further details. Later on, we show estimates that relax this assumption (Figure C2).

Figure C1: Responses to monetary policy shock from Romer and Romer (2004), with and without leads



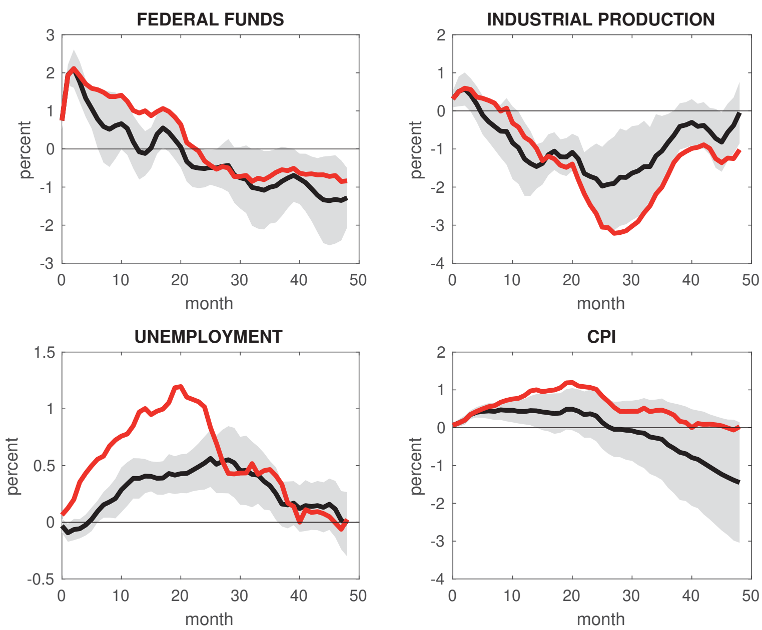
Black solid lines refer to a benchmark specification that preserves the recursive assumption and does not include leads of the shock. Grey areas show 90% Newey-West confidence intervals. Red solid lines include h leads of the monetary shocks.

The results of estimating equation (C.1) when we set $\beta_{h,f} = 0 \; \forall h, f$ are shown in solid black lines (with 90% confidence bands) in Figure C1. Since we employ the same data and specification, they replicate the results from Ramey (2016) (Figure 2B). Ramey (2016) argues that the responses using LPs show more plausible dynamics than those obtained from a standard VAR (a persistent fall in industrial output and a rise in unemployment that slowly converge to 0). The drop in output after a monetary shock is broadly consistent with the original results from Romer and Romer (2004) but there are, however, two important differences. First, the trough in the response of output is reached after two years. In the estimates of Figure C1 and Ramey (2016), the trough is reached after a first year and lasts for about twelve months with a slight rebound in between. Second, although both results refer to the same impulse (a realization of the policy measure of one percentage point), the magnitude of the output fall in the original estimates of Romer and Romer (2004) is substantially bigger than when using LPs (-4.3 *vs* -1.7).

Next, we investigate whether accounting for the persistence in the shock has an effect on the dynamic responses. We re-estimate equation (C.1), but allowing $\beta_{h,f} \neq 0$. The results are shown in red lines in Figure C1. We observe that the dynamics of output are closer to the original estimates of Romer and Romer (2004): a continuous drop in output that reaches the trough after the second year. Furthermore, the magnitude of the fall is now substantially higher (-3.1) and closer to the results from Romer and Romer (2004). Another noticeable difference is that the effects on unemployment and the initial positive reaction on prices (the so-called *price puzzle*) are now larger. All in all, the results from Figure C1 suggest that accounting for the persistence of the monetary policy shock can lead to larger estimates of the dynamic responses.

Ramey (2016) also investigates the role of the recursiveness assumption in the dynamic responses (the inclusion of contemporaneous values for some variables in the LPs estimation to replicate the identification in a VAR). She finds that relaxing this assumption results in weird dynamics of unemployment in the short run. We replicate these results by dropping the contemporaneous values in $\boldsymbol{x}_t$ in equation (C.1) and setting $\beta_{h,f} = 0$. We indeed find that unemployment rate significantly drops in the first months after a monetary policy contraction (black solid lines in Figure C2). We investigate whether these strange dynamics may be the result of the persistence in the monetary policy shock. We estimate again equation (C.1) relaxing both the recursiveness assumption and allowing $\beta_{h,f} \neq 0$. The results (red solid lines in Figure C2) are very similar to those from Figure C1. Interestingly, unemployment responds positively to the monetary policy contraction.

Figure C2: Responses to monetary policy shock from Romer and Romer (2004) with no recursive assumption, with and without leads



Black solid lines refer to a benchmark specification that relaxes the recursive assumption and does not include leads of the shock. Grey areas show 90% Newey-West confidence intervals. Red solid lines include h leads of the monetary shocks.

C.2 Gertler and Karadi (2015)

In this subsection we explore another application of the effects of monetary policy shocks based on Gertler and Karadi (2015). The authors identify exogenous changes in monetary policy by looking at variations in the 3-month-ahead futures of the federal funds within a 30-minute window of a FOMC announcement.[3] By relying on this identification scheme,

[3] This scheme is often denoted as High-Frequency Identification (HFI). See Ramey (2016) for a comparison with other identification procedures.

rather than on standard timing assumptions (e.g. Christiano et al. (1999)), the authors are able to explore the effects on measures of financial market frictions or other variables that are often assumed to be contemporaneously invariant to a monetary policy shock.

As shown in Table 1, this measure of monetary policy shock displays some persistence. This was first noted by Ramey (2016), who highlights that the procedure followed by Gertler and Karadi (2015) to convert FOMC shocks (expressed at FOMC frequency) to monthly frequency introduces this.[4]

Gertler and Karadi (2015) embedded the identified monetary policy shocks in a VAR, using the measure of monetary policy surprises as an instrument of the residuals in the VAR. Here we explore what consequences the pesistence of the shock might have if the researcher were to use LPs.

To do so, we implement the following specification, suggested by Ramey (2016):

$$y_{t+h} = \beta_{h,0} shock_t + \boldsymbol{\theta_h(L)x_t} + \beta_{h,f} \sum_{f=1}^{min\{h,12\}} shock_{t+f} + \xi_{t+h}, \tag{C.2}$$

where y_t is either the 1-year government bond rate, the log of industrial production, the excess bond premium spread from Gilchrist and Zakrajšek (2012), or the log of consumer price index, and $shock_t$ is the measure of monetary policy shocks from Gertler and Karadi (2015). The regressions also include a set of controls $\boldsymbol{x}_t$, with two lags and the contemporaneous values of all dependent variables, and two lags of the shock.[5] Following Ramey (2016), we estimate equation (C.2) for a sample of 1991:m1-2012m6.[6] Given this reduced sample, we limit the number of leads introduced in the estimation to a maximum of 12 (i.e. we use h leads for $h < 12$ and 12 leads for longer horizons).[7]

The results of these estimations are shown in Figure C3 for two cases: setting $\beta_{h,f} = 0$ $\forall h, f$ (black solid lines with 68 and 90% confidence intervals) and allowing $\beta_{h,f} \neq 0$ (red lines). Benchmark estimations that do not account for persistence reproduce the results from Ramey (2016) (Figure 3B). She notes that LPs using directly the Gertler and Karadi (2015) instrument as an explanatory variable give rise to puzzling results, namely: a sluggish response of the policy rate, output increases after the monetary expansion, and the credit

[4] In particular, Gertler and Karadi (2015) cumulate the surprises on any FOMC days during the last 31 days, effectively introducing a first-order moving-average structure. This is a variation of the procedure followed by Romer and Romer (2004) and that also results in a measure of monetary policy shocks that displays persistence.

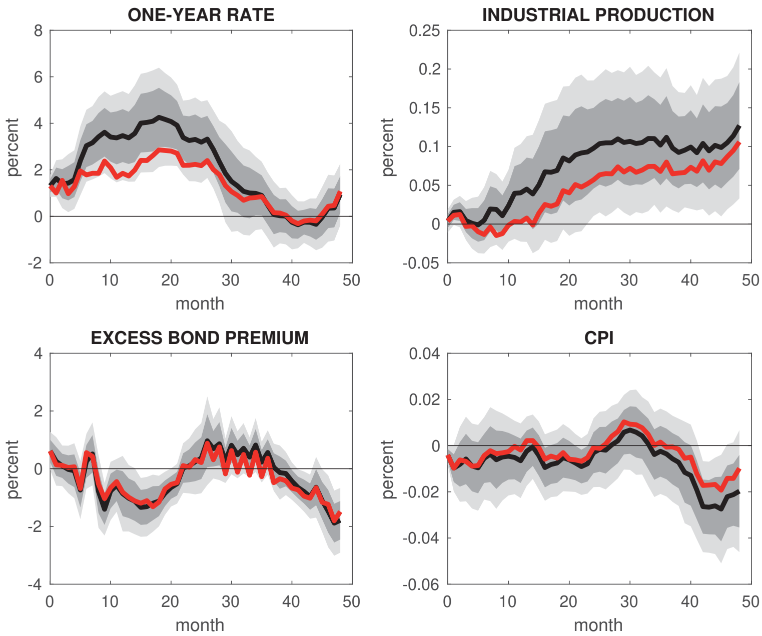
[5] Note that while the inclusion of lags of the shocks are meant to account for in the shock, our analysis from Section 3 shows that they are not effective for this role. In our results, the inclusion of lags of the shock did not have any noticeable effect. We keep them here in order to replicate the results from Ramey (2016).

[6] Gertler and Karadi (2015) proceed in two steps: they first estimate the dynamic coefficients and residuals from a VAR during the period 1979-2012. Then they estimate the contemporaneous effects of monetary policy using both the residuals from the previous step and the monetary policy instrument in a proxy VAR during 1991-2012.

[7] We also preserve the sample at the end of the period until 2012m06 (which will be otherwise reduced when including leads) by considering values of the leads of the shock equal to 0 for the last 12 periods. Although this is not important for our results, it allows us to compare our estimates to those from Ramey (2016).

spread and prices do not show significant dynamics during most of the response horizon. Ramey (2016) suggests that the persistence exhibited by Gertler and Karadi (2015) and potential predictability of the series could be sources of concern. When we incorporate leads of the shock in the estimation of equation (C.2) (red lines in Figure C3) we see that both the response of the government bond rate and industrial production seem to be overestimated when persistence is not accounted for (by contrast, the results do not change much for the excess bond premium and prices).[8]

Figure C3: Responses to monetary policy shock from Gertler and Karadi (2015), with and without leads



Black solid lines refer to a benchmark specification that does not include leads of the shock. Grey areas show 68 and 90% Newey-West confidence intervals. Red solid lines include h leads of the monetary shocks up to horizon $h = 12$, after then, the number of leads is kept to 12.

[8] Alternatively, Ramey (2016) concludes that these differences may be due to the fact that the reduced-form parameters (used to construct the impulse responses) are estimated for a longer sample (1970-2012 instead of 1991-2012) or to potential misspecification of the original VAR estimates due to the rising importance of forward guidance, which may lead to a problem of non-fundamentalness in the VAR.

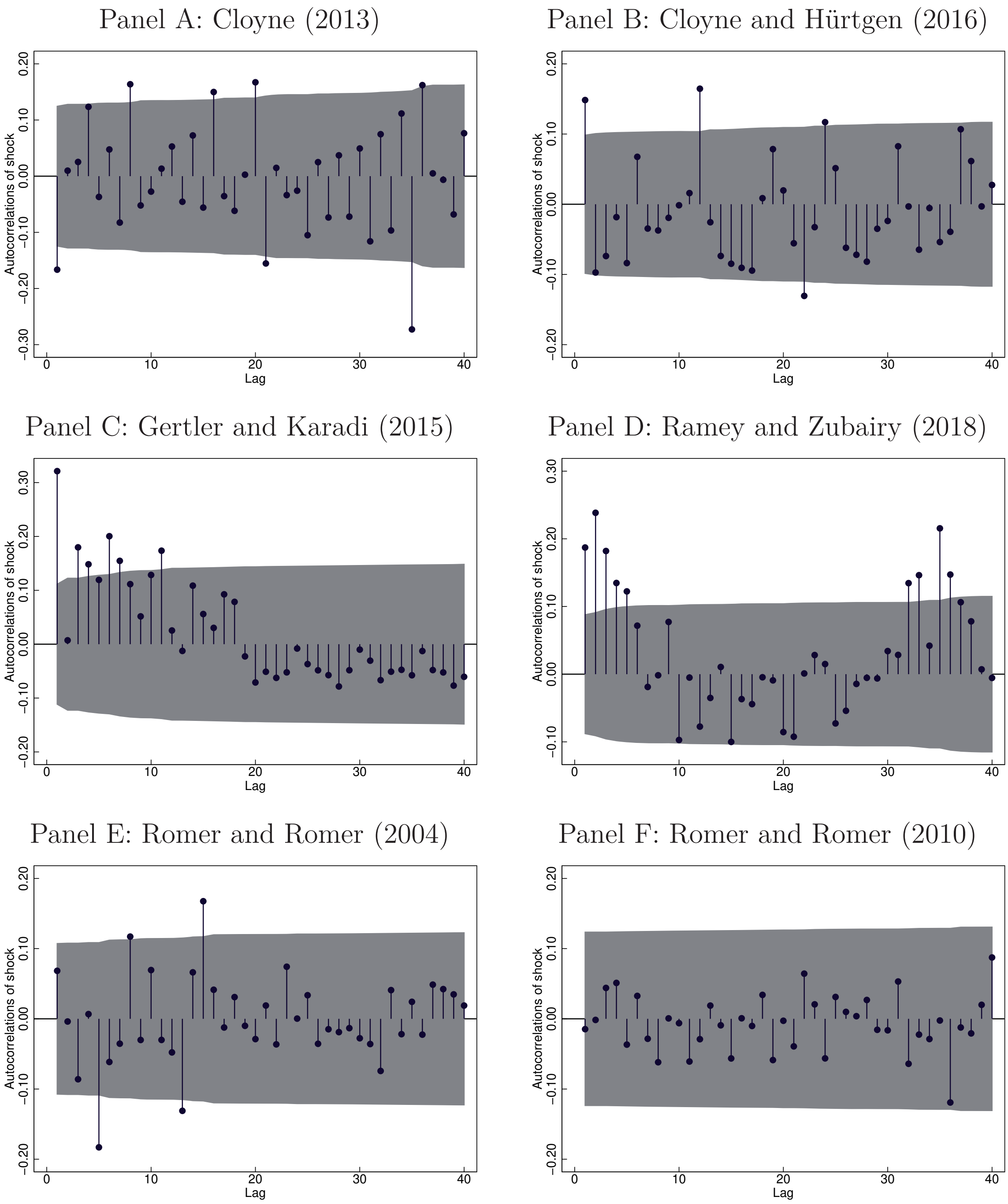
D Additional Tables and Figures

Table D.1: Robustness: different lag structures for tests

	5 lags	10 lags	20 lags	40 lags	60 lags
Arezki et al. (2017)	175.944	175.953	176.049	177.903	177.907
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cloyne (2013)	11.365	21.521	40.041	98.751	120.270
	(0.045)	(0.018)	(0.005)	(0.000)	(0.000)
Cloyne and Hürtgen (2016)	17.723	20.771	47.357	84.422	103.001
	(0.003)	(0.023)	(0.001)	(0.000)	(0.001)
Gertler and Karadi (2015)	53.802	84.284	106.133	124.568	131.030
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Guajardo et al. (2014)	160.740	173.315	182.866	185.810	185.810
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ramey and Zubairy (2018)	79.298	89.916	104.414	182.950	190.974
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Romer and Romer (2004)	15.536	23.965	43.824	53.758	64.576
	(0.008)	(0.008)	(0.002)	(0.072)	(0.320)
Romer and Romer (2010)	1.578	3.080	6.562	19.023	24.783
	(0.904)	(0.980)	(0.998)	(0.998)	(1.000)

The columns report the values of a Box and Pierce (1970) test (with Ljung and Box (1978) correction) including different lags. P-values are shown in brackets. in Arezki et al. (2017) and Guajardo et al. (2014) is tested using a generalized version of the autocorrelation test proposed by Arellano and Bond (1991) that specifies the null hypothesis of no autocorrelation at a given lag order.

Figure D1: Autocorrelograms



95% confidence intervals computed using Bartlett's formula for MA(q) processes.

Figure D2: Time series of shocks

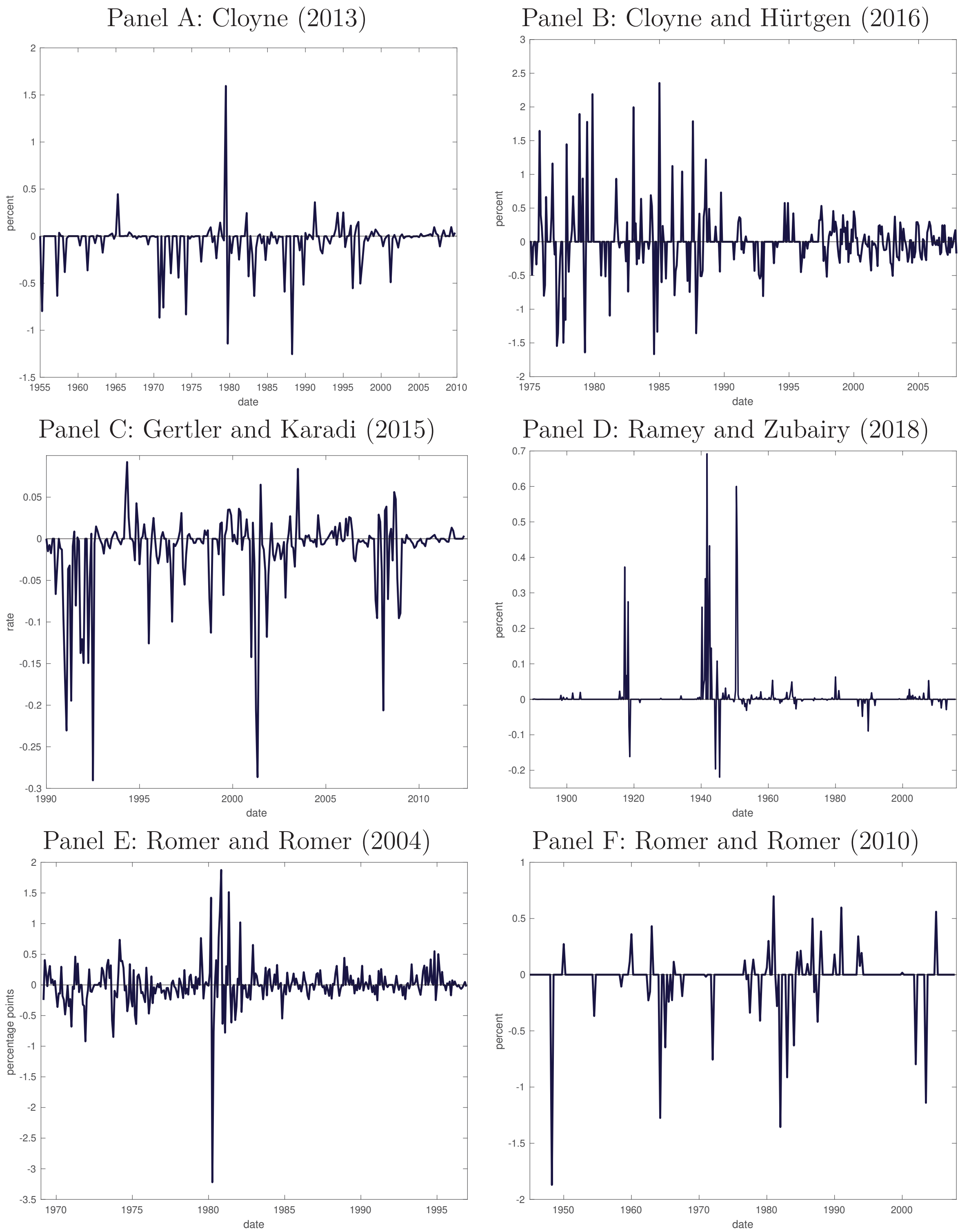
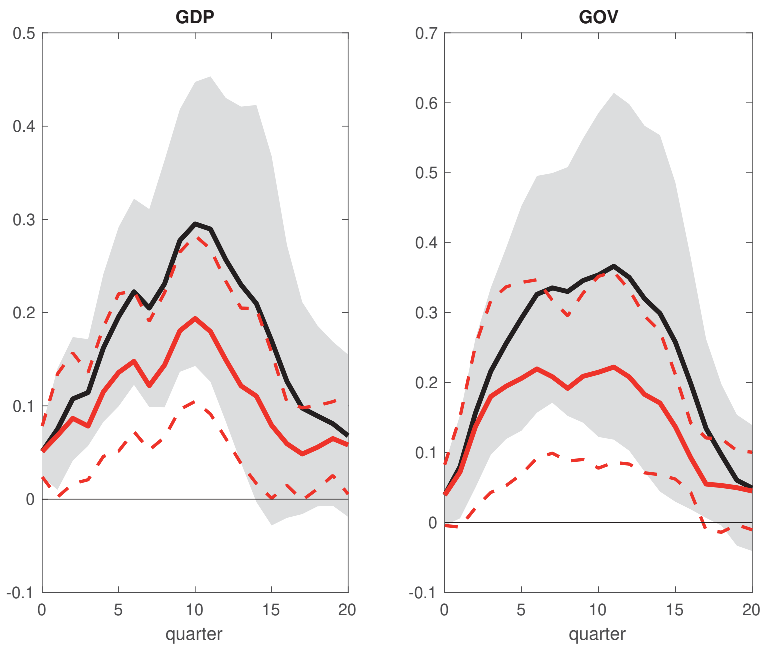
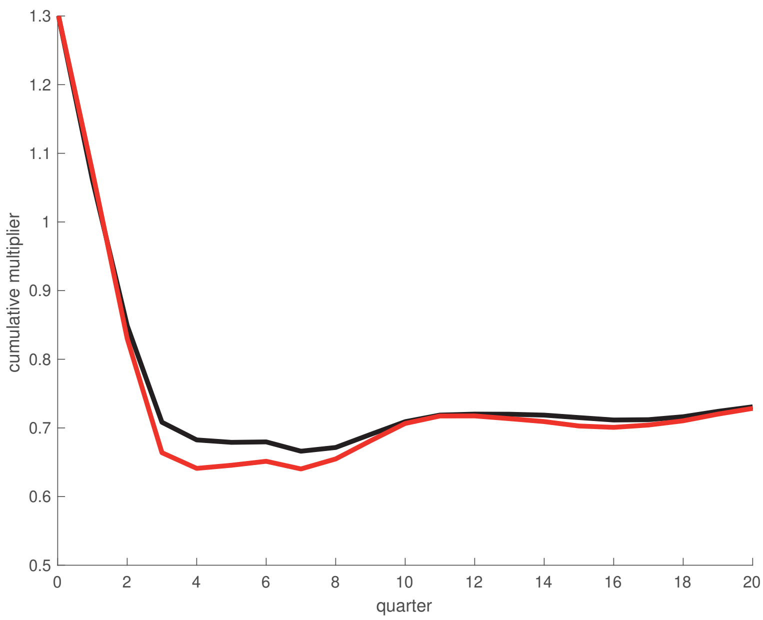


Figure D3: Output and government spending responses, with and without leads



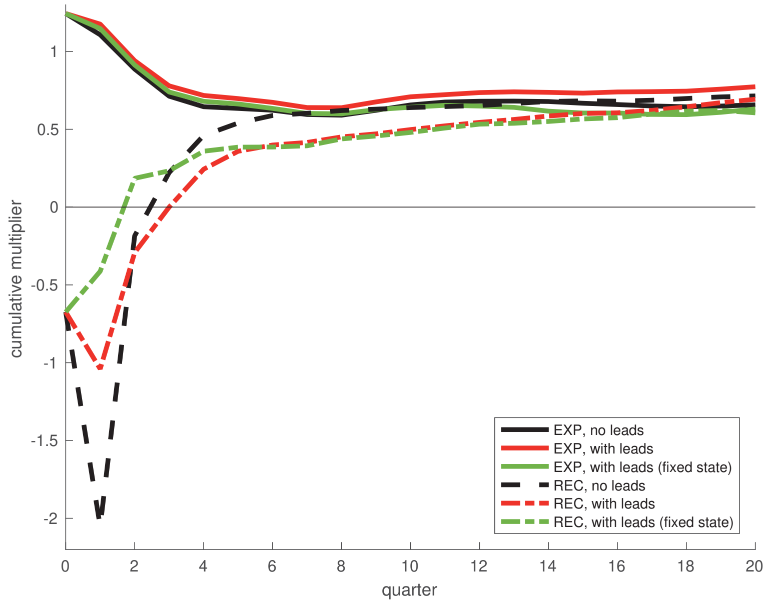
Black lines show the results of estimating the system (17) without including any lead (as in Ramey and Zubairy (2018)). Grey areas represent 68 and 95% Newey-West confidence intervals for these estimates. Red solid lines represent the results of estimations when including h leads of the Ramey and Zubairy (2018) news variable (with 95% Newey-West confidence intervals).

Figure D4: Government spending multiplier, with and without leads



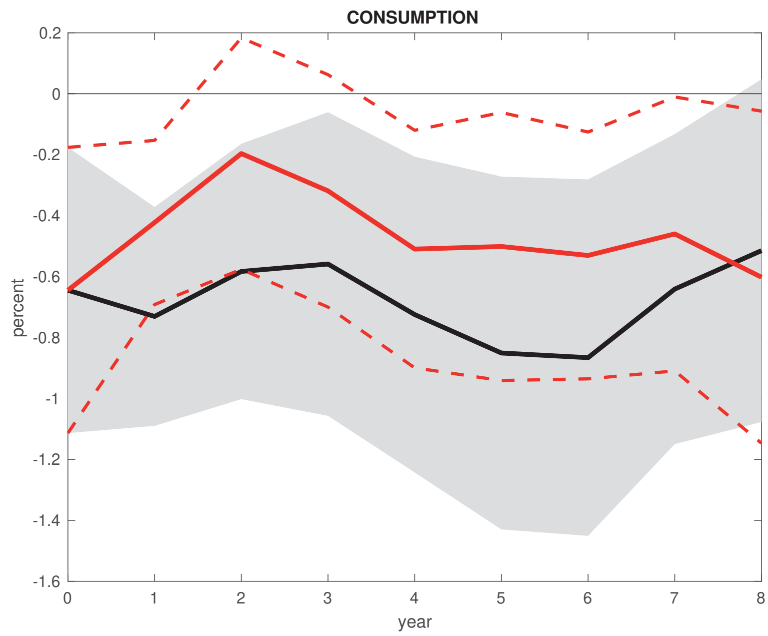
Black lines show the cumulative multiplier without including any lead. Red solid lines represent the estimates of the cumulative multiplier when including a number of leads of the Ramey and Zubairy (2018)) news variable that increase with the response horizon.

Figure D5: Government spending multiplier during expansions and recessions, with and without leads



The black solid and dashed lines show the cumulative multiplier during periods of expansion and recession, respectively, without including any lead (as in Ramey and Zubairy (2018)). The red solid and dashed lines show the cumulative multiplier during periods of expansion and recession, respectively, when including leads of the shocks and the state. Green solid and dashed lines refer to estimates of the expansion and recession multipliers, respectively, when including leads of the shock and the regime.

Figure D6: Private consumption response to a fiscal consolidation shock, with and without leads



Black lines show the results from equation (20) with private consumption as dependent variable and setting $\beta_{h,f} = 0$, i.e. without including any lead of the shock. Grey areas represent 90% Newey-West confidence intervals for these estimates (save interval as reported in Guajardo et al. (2014)). Red solid lines represent the results of estimations when allowing $\beta_{h,f} \neq 0$ and including h leads of the consolidations variable.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

1840 ALESSIO MORO and OMAR RACHEDI: The changing structure of government consumption spending.

1841 GERGELY GANICS, ATSUSHI INOUE and BARBARA ROSSI: Confidence intervals for bias and size distortion in IV and local projections – IV models.

1842 MARÍA GIL, JAVIER J. PÉREZ, A. JESÚS SÁNCHEZ and ALBERTO URTASUN: Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data.

1843 MATÍAS LAMAS and JAVIER MENCÍA: What drives sovereign debt portfolios of banks in a crisis context?

1844 MIGUEL ALMUNIA, POL ANTRÀS, DAVID LÓPEZ-RODRÍGUEZ and EDUARDO MORALES: Venting out: exports during a domestic slump.

1845 LUCA FORNARO and FEDERICA ROMEI: The paradox of global thrift.

1846 JUAN S. MORA-SANGUINETTI and MARTA MARTÍNEZ-MATUTE: An economic analysis of court fees: evidence from the Spanish civil jurisdiction.

1847 MIKEL BEDAYO, ÁNGEL ESTRADA and JESÚS SAURINA: Bank capital, lending booms, and busts. Evidence from Spain in the last 150 years.

1848 DANIEL DEJUÁN and CORINNA GHIRELLI: Policy uncertainty and investment in Spain.

1849 CRISTINA BARCELÓ and ERNESTO VILLANUEVA: The risk of job loss, household formation and housing demand: evidence from differences in severance payments.

1850 FEDERICO TAGLIATI: Welfare effects of an in-kind transfer program: evidence from Mexico.

1851 ÓSCAR ARCE, GALO NUÑO, DOMINIK THALER and CARLOS THOMAS: A large central bank balance sheet? Floor vs corridor systems in a New Keynesian environment.

1901 EDUARDO GUTIÉRREZ and ENRIQUE MORAL-BENITO: Trade and credit: revisiting the evidence.

1902 LAURENT CAVENAILE and PAU ROLDAN: Advertising, innovation and economic growth.

1903 DESISLAVA C. ANDREEVA and MIGUEL GARCÍA-POSADA: The impact of the ECB's targeted long-term refinancing operations on banks' lending policies: the role of competition.

1904 ANDREA ALBANESE, CORINNA GHIRELLI and MATTEO PICCHIO: Timed to say goodbye: does unemployment benefit eligibility affect worker layoffs?

1905 CORINNA GHIRELLI, MARÍA GIL, JAVIER J. PÉREZ and ALBERTO URTASUN: Measuring economic and economic policy uncertainty, and their macroeconomic effects: the case of Spain.

1906 CORINNA GHIRELLI, JAVIER J. PÉREZ and ALBERTO URTASUN: A new economic policy uncertainty index for Spain.

1907 ESTEBAN GARCÍA-MIRALLES, NEZIH GUNER and ROBERTO RAMOS: The Spanish personal income tax: facts and parametric estimates.

1908 SERGIO MAYORDOMO and OMAR RACHEDI: The China syndrome affects banks: the credit supply channel of foreign import competition.

1909 MÓNICA CORREA-LÓPEZ, MATÍAS PACCE and KATHI SCHLEPPER: Exploring trend inflation dynamics in Euro Area countries.

1910 JAMES COSTAIN, ANTON NAKOV and BORJA PETIT: Monetary policy implications of state-dependent prices and wages.

1911 JAMES CLOYNE, CLODOMIRO FERREIRA, MAREN FROEMEL and PAOLO SURICO: Monetary policy, corporate finance and investment.

1912 CHRISTIAN CASTRO and JORGE E. GALÁN: Drivers of productivity in the Spanish banking sector: recent evidence.

1913 SUSANA PÁRRAGA RODRÍGUEZ: The effects of pension-related policies on household spending.

1914 MÁXIMO CAMACHO, MARÍA DOLORES GADEA and ANA GÓMEZ LOSCOS: A new approach to dating the reference cycle.

1915 LAURA HOSPIDO, LUC LAEVEN and ANA LAMO: The gender promotion gap: evidence from Central Banking.

1916 PABLO AGUILAR, STEPHAN FAHR, EDDIE GERBA and SAMUEL HURTADO: Quest for robust optimal macroprudential policy.

1917 CARMEN BROTO and MATÍAS LAMAS: Is market liquidity less resilient after the financial crisis? Evidence for US treasuries.

1918 LAURA HOSPIDO and CARLOS SANZ: Gender Gaps in the Evaluation of Research: Evidence from Submissions to Economics Conferences.

1919 SAKI BIGIO, GALO NUÑO and JUAN PASSADORE: A framework for debt-maturity management.

1920 LUIS J. ÁLVAREZ, MARÍA DOLORES GADEA and ANA GÓMEZ-LOSCOS: Inflation interdependence in advanced economies.

1921 DIEGO BODAS, JUAN R. GARCÍA LÓPEZ, JUAN MURILLO ARIAS, MATÍAS J. PACCE, TOMASA RODRIGO LÓPEZ, JUAN DE DIOS ROMERO PALOP, PEP RUIZ DE AGUIRRE, CAMILO A. ULLOA and HERIBERT VALERO LAPAZ: Measuring retail trade using card transactional data.

1922 MARIO ALLOZA and CARLOS SANZ: Jobs multipliers: evidence from a large fiscal stimulus in Spain.

1923 KATARZYNA BUDNIK, MASSIMILIANO AFFINITO, GAIA BARBIC, SAIFFEDINE BEN HADJ, ÉDOUARD CHRÉTIEN, HANS DEWACHTER, CLARA ISABEL GONZÁLEZ, JENNY HU, LAURI JANTUNEN, RAMONA JIMBOREAN, OTSO MANNINEN, RICARDO MARTINHO, JAVIER MENCÍA, ELENA MOUSARRI, LAURYNAS NARUŠEVIČIUS, GIULIO NICOLETTI, MICHAEL O'GRADY, SELCUK OZSAHIN, ANA REGINA PEREIRA, JAIRO RIVERA-ROZO, CONSTANTINOS TRIKOUPIS, FABRIZIO VENDITTI and SOFÍA VELASCO: The benefits and costs of adjusting bank capitalisation: evidence from Euro Area countries.

1924 MIGUEL ALMUNIA and DAVID LÓPEZ-RODRÍGUEZ: The elasticity of taxable income in Spain: 1999-2014.

1925 DANILO LEIVA-LEON and LORENZO DUCTOR: Fluctuations in global macro volatility.

1926 JEF BOECKX, MAARTEN DOSSCHE, ALESSANDRO GALESI, BORIS HOFMANN and GERT PEERSMAN: Do SVARs with sign restrictions not identify unconventional monetary policy shocks?

1927 DANIEL DEJUÁN and JUAN S. MORA-SANGUINETTI: Quality of enforcement and investment decisions. Firm-level evidence from Spain.

1928 MARIO IZQUIERDO, ENRIQUE MORAL-BENITO and ELVIRA PRADES: Propagation of sector-specific shocks within Spain and other countries.

1929 MIGUEL CASARES, LUCA DEIDDA and JOSÉ E. GALDÓN-SÁNCHEZ: On financial frictions and firm market power.

1930 MICHAEL FUNKE, DANILO LEIVA-LEON and ANDREW TSANG: Mapping China's time-varying house price landscape.

1931 JORGE E. GALÁN and MATÍAS LAMAS: Beyond the LTV ratio: new macroprudential lessons from Spain.

1932 JACOPO TIMINI: Staying dry on Spanish wine: the rejection of the 1905 Spanish-Italian trade agreement.

1933 TERESA SASTRE and LAURA HERAS RECUERO: Domestic and foreign investment in advanced economies. The role of industry integration.

1934 DANILO LEIVA-LEON, JAIME MARTÍNEZ-MARTÍN and EVA ORTEGA: Exchange rate shocks and inflation comovement in the euro area.

1935 FEDERICO TAGLIATI: Child labor under cash and in-kind transfers: evidence from rural Mexico.

1936 ALBERTO FUERTES: External adjustment with a common currency: the case of the euro area.

1937 LAURA HERAS RECUERO and ROBERTO PASCUAL GONZÁLEZ: Economic growth, institutional quality and financial development in middle-income countries.

1938 SILVIA ALBRIZIO, SANGYUP CHOI, DAVIDE FURCERI and CHANSIK YOON: International Bank Lending Channel of Monetary Policy.

1939 MAR DELGADO-TÉLLEZ, ENRIQUE MORAL-BENITO and JAVIER J. PÉREZ: Outsourcing and public expenditure: an aggregate perspective with regional data.

1940 MYROSLAV PIDKUYKO: Heterogeneous spillovers of housing credit policy.

1941 LAURA ÁLVAREZ ROMÁN and MIGUEL GARCÍA-POSADA GÓMEZ: Modelling regional housing prices in Spain.

1942 STÉPHANE DÉES and ALESSANDRO GALESI: The Global Financial Cycle and US monetary policy in an interconnected world.

1943 ANDRÉS EROSA and BEATRIZ GONZÁLEZ: Taxation and the life cycle of firms.

1944 MARIO ALLOZA, JESÚS GONZALO and CARLOS SANZ: Dynamic effects of persistent shocks.

BANCO DE ESPAÑA
Eurosistema

Unidad de Servicios Auxiliares
Alcalá, 48 - 28014 Madrid
E-mail: publicaciones@bde.es
www.bde.es