Dynamic Effects of Persistent Shocks

Mario Alloza* Jesús Gonzalo* Carlos Sanz*

Bank of Spain Universidad Carlos III de Madrid Bank of Spain and CEMFI

October 29, 2023

Abstract

We provide evidence that many narrative shocks used by prominent literature display some persistence. We show that the two leading methods to estimate impulse responses to an independently identified shock (local projections and distributed lag models) treat persistence differently, hence identifying different objects. We propose corrections to re-establish the equivalence between local projections and distributed lag models, providing applied researchers with methods and guidance to estimate their desired object of interest. We apply these methods to well-known empirical work and find that how persistence is treated 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.

*We thank Fabio Canova, Jesús Fernández-Villaverde, Alessandro Galesi, Gergely Gánics, Juan F. Jimeno, Evi Pappa, Gabriel Pérez-Quirós, Mikkel Plagborg-Møller, Valerie Ramey, Juan Rubio-Ramírez, Enrique Sentana, and seminar participants at the I Workshop of the Spanish Macroeconomics Network (Universidad Pública de Navarra), Bank of Spain, CFE 2018 (University of Pisa), IAAE 2019 (Cyprus), 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. Gonzalo gratefully acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación through grants P!D2019-104960GB-IO0, TED2021-129784B-IO0, AEl/10.13039/501100011033-CEX2021-001181-M.

Alloza: m.alloza@bde.es. Gonzalo: jesus.gonzalo@uc3m.es. Sanz: carlossanz@bde.es.

1 Introduction

Estimating the impact of economic shocks is a crucial aspect of macroeconomics. To identify economically meaningful shocks, the literature has traditionally relied on systems of equations coupled with restrictions implied by economic theory. Recently, researchers are increasingly using narrative identification, e.g., looking at written official documentation or newspapers and exploiting arguably exogenous variation in these series— see Romer and Romer (2004), Romer and Romer (2010), or Ramey and Zubairy (2018) for prominent examples of narrative identification. While its focus on identifying exogenous variation is appealing, the lack of restrictions in narrative methods yields objects with less standard time series properties.

In this paper, we analyze how the presence of persistence in narrative shocks affects the identification and estimation of their dynamic effects, providing empirical researchers with methods and guidance to deal with this issue.[1]

We begin by showing that many narrative shocks used by prominent literature are serially correlated. In particular, we systematically test for serial correlation in eight shocks used in leading economics journals. We find evidence of serial correlation in seven of them. The presence of persistence in the shock does not necessarily preclude these variables from being categorized as "shocks" following standard definitions of aggregate shocks. More concretely, according to Ramey (2016), a shock should represent unanticipated movements. What this condition implies is that shocks are unforecastable, i.e., they are forecast errors. In particular, when the forecasting loss function is not quadratic (for instance, the check function), the forecasting errors may not be a martingale difference sequence (m.d.s) and therefore could be serially correlated. However, serial correlation poses additional challenges for the identification of the macroeconomic experiment of interest.

When estimating the dynamic response of some variable to a serially correlated shock, some part of this persistence may be passed on to the impulse response function (IRF).

[1] Throughout the paper we use the term *persistence* as a phenomenon captured or reflected by *serial correlation*, a testable condition. We use both terms interchangeably.

Hence, a researcher may want to identify one of two objects of interest: the response *as if* the shock were uncorrelated, i.e., to a counter-factual serially uncorrelated shock ($\mathcal{R}(h)^*$), or the response to the shock as it is, i.e., including the effect of persistence in the IRF ($\mathcal{R}(h)$). Deciding for one or the other depends on what specific question the researcher is trying to address. On the one hand, $\mathcal{R}(h)^*$ allows to compare effects with those obtained from a theoretical or empirical model, and facilitates comparisons across different types of shocks (e.g., monetary versus fiscal shocks) or across countries. On the other hand, $\mathcal{R}(h)$ is more appropriate if the researcher is interested in evaluating the *most likely dynamic response* of a variable to a shock based on historical data. Regardless of which object is preferred by the researcher, the difference between $\mathcal{R}(h)$ and $\mathcal{R}(h)^*$ is informative about how much of the dynamic transmission of a shock is due to the presence of persistence.

We consider the two most popular methods to estimate impulse responses when a shock has already been identified (e.g., using narrative methods). These are local projections (LPs) (Jordà (2005)) and distributed lag models (DLMs). By DLMs we refer to single-equation regressions of an outcome variable against the contemporaneous value and lags of the shock with or without an autoregressive component, as in Romer and Romer (2004), Cerra and Saxena (2008), Romer and Romer (2010), Alesina et al. (2015), Arezki et al. (2017), and Coibion et al. (2018). We show that, if there is no serial correlation, the two methods identify the same object. However, we demonstrate that this equivalence breaks down in the presence of serial correlation. In this case, LPs identify $\mathcal{R}(h)$ while DLMs regressions identify $\mathcal{R}(h)^*$. The intuition is that LPs compute the response at horizon h by regressing the outcome variable in $t+h$ against the shock in time t. Since the standard setting does not account for how the shock evolves between t and $t+h$, the responses include 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. By contrast, DLMs implicitly account for the evolution of the shock, hence identifying the effect as if the shock were not persistent.

While this result might seem discouraging for empirical work, we then show that it is possible to adjust both estimating methods to obtain the desired object of interest. Consider a researcher who wants to use LPs and is interested in identifying $\mathcal{R}(h)^*$. As mentioned, if she runs standard LPs with a persistent shock, she will identify $\mathcal{R}(h)$ instead. Perhaps surprisingly, the most obvious solution of including lags of the shock will not address this issue. However, we show that, by including *leads* of the shock, she will recover $\mathcal{R}(h)^*$. Likewise, we show how standard DLMs can be adapted so that they identify $\mathcal{R}(h)$.

We also show that the lack of equivalence between LPs and DLMs under persistence carries over to multivariate settings. In particular, the equivalence between VARs with the shock embedded as an endogenous vs. an exogenous variable breaks down in the presence of serial correlation, which has relevant practical implications for applied researchers.[2] Our analysis of multivariate settings opens up new avenues, like using LPs to uncover the dynamic relations of two variables by including leads of a third variable, hence obtaining the response *as if* this third variable had remained constant over the response horizon. This can be seen as the LP counterpart 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 as in Bernanke et al. (1997), Sims and Zha (2006), or Bachmann and Sims (2012).

To illustrate how our methods work, we consider an empirical application, which also serves to assess the quantitative relevance of persistence in a real case by comparing estimates of $\mathcal{R}(h)$ and $\mathcal{R}(h)^*$. In particular, we consider Ramey and Zubairy (2018)'s LPs estimation of the dynamic effects to a shock constructed from news about future changes in defense spending. We find that, after two years, the responses that exclude the effect of persistence in the shock are about 40% lower than the original Ramey and Zubairy (2018)'s estimates. The effect of serial correlation also seems to have an effect on the short-run response of fiscal multipliers during recessions.

[2] Note that the equivalence arises because a VAR with a shock as an exogenous variable (often known as VAR-X) can be seen as multivariate generalization of a DLM (see Mertens and Ravn (2012) or Favero and Giavazzi (2012) for examples of VAR-X specifications).

In the appendix, we consider two additional applications, based on Guajardo et al. (2014) and Romer and Romer (2010). Overall, we find that how persistence is treated can have a sizable impact on the estimated effects. Hence, computing the statistic $\mathcal{R}(h) - \mathcal{R}(h)^*$ can help the empirical researcher to recognise how much of the dynamic response of a variable depends on the serial correlation of a shock.

Our paper makes four contributions to the literature. First, we formally and systematically test for the presence of serial correlation in shocks used by previous work. Although the issue of persistence in shocks has been noted before,[3] we believe we are the first to formally and systematically test for serial correlation in prominent narratively-identified shocks.

Our second contribution is to show that, while both LPs and DLMs identify the same object if the shock is serially uncorrelated, this equivalence breaks down in the presence of persistence. Plagborg-Møller and Wolf (2021) prove that LPs and VAR methods identify the same impulse responses when both methods have an unrestricted lag structure. This result formalizes some of the examples provided in Ramey (2016), which implies that different identification schemes in a VAR setting can be implemented in a LP context. Our result builds on a different premise: we consider the cases where the shock has already been identified using narrative measures and the researcher wants to use LPs or DLMs to estimate dynamic effects.

Our third contribution is to provide methods to re-establish the LP-DLM equivalence when there is persistence, providing applied researchers with a menu of options to identify their desired object of interest. In this regard, our method of adding leads to LPs is related to the tradition in factor analysis by Geweke and Singleton (1981) and on the DOLS estimation of cointegration vectors (Stock and Watson (1993)). Dufour and Renault (1998) introduce leads in some of their IRFs to study causality at different horizons. 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

[3] 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.

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 DLMs that we identify is not due to a bias in the estimates, but instead to differences in identification due to the persistence of the shock. Since our problem still persists asymptotically, increasing the sample does not reduce the LP-DLM difference. Additionally, this difference is not necessarily negative, but will depend on the nature of the data generating process (DGP) that drives the persistence.

Finally, we speak to some recent and well-known empirical work on the effects of monetary and fiscal policy (Ramey and Zubairy (2018) Guajardo et al. (2014), Romer and Romer (2004), and Gertler and Karadi (2015)). Our contribution is to apply our methods to these works and re-assess their empirical evidence. We do not claim that any of these papers is "wrong". Rather, what our results suggest is that the interpretation of their results depends on the desired object of interest and the employed estimating method.

The rest of the paper proceeds as follows. Section 2 provides evidence of serial correlation in shocks used by previous work. Section 3 shows that LPs and DLMs treat persistence differently, and proposes a solution to re-establish the equivalence between them. It also introduces generalizations to VARs and discusses briefly the case of LP-IV. Section 4 lays out an application based on Ramey and Zubairy (2018). Section 5 concludes. The appendix contains proofs of the theoretical results, simulations, and additional empirical applications.

2 Evidence and implications of serial correlation in 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. Typically, this implies that the resulting shocks are well-behaved and display some statistical features that might be seen as desirable—in particular, no persistence. Alternatively, shocks may be identified without the use of a model, for example, by using *narrative* methods. This alternative identification relies on the existence of historical sources, such as official documentation, periodicals, etc., from which a shock variable is constructed. In this section, we provide evidence that it is common that shocks identified this way are persistent. We then take stock on this finding in light of Ramey (2016)'s canonical definition of a shock.

We study eight aggregate shocks used by prominent literature on 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.[4] The literature has used these variables mainly as direct measures of shocks, although it sometimes refers to these shocks as instruments or proxies.

To test for the presence of persistence we use a *portmanteau*-type test following Box and Pierce (1970). 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. The null hypothesis is that the data are not serially correlated. We test for

[4] 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 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 orthogonalize 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. Note that the present analysis could also be generalized to other type of innovations obtained without the use of a model such as climate shocks (defined as deviations of the temperatures from the their corresponding trend), which might also be serially correlated.

the presence of autocorrelation in 40 periods. 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 uncorrelation 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%.[5] 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.[6]

According to the canonical definition (Ramey (2016)), empirical shocks should (i) be exogenous to current and lagged endogenous variables, (ii) be uncorrelated to other exogenous shocks, and (iii) represent unanticipated movements (or news about future shocks). While one might think that the presence of persistence violates the third condition, this is not necessarily the case. When the forecasting loss function is the quadratic one, it is well known that the forecasting errors must be a m.d.s with respect to some information set and therefore uncorrelated—see Granger and Machina (2006) and Lee (2008) for a description and analysis of loss functions. This is the case when the shocks come directly from a conditional expectation model, like a VAR model. When the forecasting loss function is not quadratic, for instance, the check function (popular in quantile regressions), the forecasting errors are not a m.d.s and therefore they could be serially correlated. They still are forecasting errors

[5] 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.

[6] 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 they cluster around events like wars (as in Ramey and Zubairy (2018)).

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) use extended data from Ramey (2011).

(satisfy (iii)) but are serially correlated.

This indicates that serially-correlated shocks can still be labeled “shocks” according to the previous definition. However, even if a researcher always operates under the quadratic loss function and considers that serially-correlated shocks should not be called “shocks”, in the rest of the paper we show that such shocks can still provide valuable information for empirical analysis.

3 Theoretical framework

3.1 General setup

We consider the following VAR as the DGP:

$$
\begin{aligned}
\boldsymbol{y_t} &= \sum_{\ell=1}^{\infty} \boldsymbol{A_\ell y_{t-\ell}} + \sum_{q=0}^{\infty} \boldsymbol{\delta_q} x_{t-q} + \boldsymbol{u_t} \\
x_t &= \sum_{r=1}^{\infty} \gamma_r x_{t-r} + \varepsilon_t,
\end{aligned} \tag{1}
$$

where $\boldsymbol{y_t}$ is a vector of endogenous time series, all the roots of the determinant $|\boldsymbol{A(L)}| = \left|\left(\boldsymbol{I} - \sum_{\ell=1}^{\infty} \boldsymbol{A_\ell} L^\ell\right)\right|$ and the polynominal $(1-\sum_{r=1}^{\infty}\gamma_r L^r)$ are outside the unit-root circle, x_t is a strictly exogenous variable such that $\mathbb{E}\left(\boldsymbol{u_t}|x_{t-p}\right) = 0 \; \forall p \gtreqless 0$, and $\boldsymbol{u_t}$ and ε_t are a vector and a scalar of martingale difference sequences (m. d. s.) with respect to Ω_{t-1} (the information set at time t-1 containing the information available up to time t-1). They are independent from each other and with mean and variance given by $\boldsymbol{u_t} \sim (\boldsymbol{0}, \boldsymbol{\Sigma^2_u})$ and $\varepsilon_t \sim (0, \sigma^2_\varepsilon)$, respectively. Following the evidence discussed in the previous section, x_t is considered to be a shock identified using narrative methods and is allowed to be persistent. This general framework encompasses several empirical specification often found in the literature. For example, when ignoring the second equation, system (1) becomes a VAR with an exogenous variable (or VAR-X)— see, for example, Mertens and Ravn (2012) or Favero and Giavazzi (2012), which assume ℓ and q are finite numbers. Additionally, when $\boldsymbol{y_t}$ is a scalar and $\boldsymbol{A_\ell} = \boldsymbol{0}$

$\forall \ell$, system (1) becomes a DLM (as in Romer and Romer (2004) or Romer and Romer (2010)). Alternatively, when x_t is instead included in the vector of endogenous variables $\boldsymbol{y_t}$, system (1) becomes a standard VAR (as in Bloom (2009) or Ramey (2011)). We explore the implications of this last representation in Section 3.6.

Without loss of generality, we consider a simpler version of system (1), where $\boldsymbol{y_t}$ is a scalar and $\boldsymbol{A}_\ell = \boldsymbol{0}\ \forall \ell > 1$, $\boldsymbol{\delta}_q = 0\ \forall q > 1$ and $\gamma_r = 0\ \forall r > 1$:

$$\begin{aligned} y_t &= \rho y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + u_t \\ x_t &= \gamma x_{t-1} + \varepsilon_t, \end{aligned} \tag{2}$$

where y_t is now the economic outcome variable of interest (for example, growth rate of GDP), x_t is an economic shock (e.g., a fiscal or monetary policy shock) which is strictly exogenous $\mathbb{E}\left(u_t | x_{t-p}\right) = 0\ \forall p \gtreqless 0$, and u_t and ε_t are m.d.s. and independent variables with mean and variance given by $u_t \sim (0, \sigma_u^2)$ and $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$, respectively. δ_0 measures the contemporaneous impact of variable x_t on y_t and is the main parameter of interest.

The DGP described by system (2) 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. But it is sufficiently rich as to incorporate relevant empirical features such as persistence in the dependent variable and lagged effects of the shock. Importantly, the results also arise in more complex settings where we incorporate more general characteristics as in system (1), such as a multivariate setting.[7]

[7] For example, in Sections 3.6 and B.1 we consider models that also include persistence in the dependent variable and lagged effects of the shock. Appendix B.2 provides an alternative specification where the degree of serial correlation in the shock x_t is taken from real-world data.

3.2 Impulse response functions

We are interested in recovering the response of our variable of interest y_t when a shock x_t hits the system in period t. We consider two different IRFs.[8] The first, denoted by $\mathcal{R}(h)$ for period $h \geq 0$, is:

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

Importantly, note that the above definition does not condition on future realizations of x_t. Hence, if $\gamma \neq 0$, an initial unit impulse in x_t does not imply that $x_{t+j} = 0$:

$$\begin{aligned}
\mathcal{R}(0) &= \frac{\partial y_t}{\partial x_t} = \delta_0 \\
\mathcal{R}(1) &= \frac{\partial y_{t+1}}{\partial x_t} = \rho\delta_0 + \delta_0\gamma + \delta_1 \\
\mathcal{R}(2) &= \frac{\partial y_{t+2}}{\partial x_t} = \rho^2\delta_0 + \rho(\delta_0\gamma + \delta_1) + \delta_0\gamma^2 + \delta_1\gamma \\
& \dots
\end{aligned}$$

Second, the researcher might also be interested in the response to the shock *as if* the shock had no persistence. We call this second IRF $\mathcal{R}(h)^*$ and define it as the "traditional impulse response function" by Koop et al. (1996) when x_t is the shock variable of interest:

$$\mathcal{R}(h)^* = \mathbb{E}\left[y_{t+h}|x_t = 1, x_{t+1} = 0, ..., x_{t+h} = 0, \Omega_{t-1}\right] - \mathbb{E}\left[y_{t+h}|x_t = 0, x_{t+1} = 0, ..., x_{t+h} = 0, \Omega_{t-1}\right]. \tag{4}$$

It provides an answer to the question "what is the effect of a shock of size 1 hitting the system at time t on the state of the system at time $t + h$ given that no other shocks hit the system?".

Contrary to $\mathcal{R}(h)$, $\mathcal{R}(h)^*$ explicitly 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

[8] Forecast error variance decompositions (FEVD) are additional tools to analyze the dynamic effects of shocks. See Gorodnichenko and Lee (2020) for an FEVD estimator based on LPs.

of γ), i.e., the responses are observationally equivalent to those that would arise from a DGP with $\gamma = 0$: [9]

$$\begin{aligned}
\mathcal{R}(0)^* &= \frac{\partial y_t}{\partial x_t} = \delta_0 \\
\mathcal{R}(1)^* &= \frac{\partial y_{t+1}}{\partial x_t}\Big|_{x_{t+1}} = \rho\delta_0 + \delta_1 \\
\mathcal{R}(2)^* &= \frac{\partial y_{t+2}}{\partial x_t}\Big|_{x_{t+1},x_{t+2}} = \rho^2\delta_0 + \rho\delta_1 \\
& \dots
\end{aligned}$$

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 $\mathcal{R}(h) \neq \mathcal{R}(h)^* \; \forall \; h > 0$. Note that these differences are particularly visible when further assuming that $\rho = \delta_1 = 0$. In this special case, $\mathcal{R}(h)^* = 0 \; \forall \; h > 0$ while $\mathcal{R}(h) \neq 0 \; \forall \; h > 0$ if $\gamma \neq 0$.

An important question is what are the relative features of $\mathcal{R}(h)^*$ vs. $\mathcal{R}(h)$. Since $\mathcal{R}(h)^*$ can be understood as the IRF resulting from a shock that is not serially correlated, it should be the desired object when the researcher wants to establish comparisons across dynamic responses. There are at least two instances when $\mathcal{R}(h)^*$ can facilitate comparisons. First, a shock identified from within a model (say, a structural VAR) or the innovation to a stochastic process in a DSGE model are, by construction, a m.d.s. (they are non-persistent). Given the absence of serial correlation, the thought experiment carried out in such cases is equivalent to constructing and IRF such as the shock takes the value of 1 on impact and 0 afterwards. Contrary to VAR-identified shocks or innovations in a DSGE model, narratively-identified shocks may display serial correlation. If this is the case, $\mathcal{R}(h)$ will identify a different object, since the effect of serial correlation is included in the IRFs.

Second, $\mathcal{R}(h)^*$ can also be an object of interest when the researcher wants to compare the effects of different shocks, e.g., whether fiscal or monetary policy is more effective in

[9] Note that this would also correspond to the average treatment effect typically discussed in the microeconomics literature. For a link between this literature and the LPs, see Dube et al. (2023).

stimulating output.[10] For example, it may be the case that fiscal shocks tend to show more persistence or that a given identification procedure tends to generate shocks with different degree of serial correlation. If the effect of persistence amounts to a non-negligible amount of the dynamic response, this could wrongly lead to the conclusion that one shock is more effective than the other when the true underlying cause is that the DGP of both shocks is different. Since $\mathcal{R}(h)^*$ effectively computes responses to a shock that is not serially correlated, regardless of the underlying GDP, this would facilitate such comparison.

On the other hand, $\mathcal{R}(h)$ should be the object of interest when the researcher is interested in estimating the *most likely dynamic response* of a variable to a shock according to the historical data. 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 policies. If we consider the effects of a monetary policy shock that cuts the policy rate by one 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 non-persistent.

Importantly, and regardless of the experiment that one wants to run, looking at the difference between $\mathcal{R}(h)$ and $\mathcal{R}(h)^*$ is informative by itself, as it speaks about how much of the dynamic response is due to the implied DGP of the shock variable. Put differently, it informs the researcher of a propagation mechanism: $\mathcal{R}(h)$ includes the propagation through the persistence of x_t while $\mathcal{R}(h)^*$ does not. Hence, the statistic $\mathcal{R}(h)-\mathcal{R}(h)^*$ can be an useful addition to the empirical researcher's toolbox to explore how the properties of an identified

[10] This also applies to a comparison of responses to the same shock with, say, data from different countries. Consider the following example: we want to compare the effects of fiscal policy in the US (using a news variable) and in another country (where we have availability of an alternative news variables). Consider the case that the news variables have different amounts of serial correlation and we obtain estimates of the government spending multipliers in both countries. Could we conclude that government spending is more effective in one country versus the other? Potentially, both policies could be equally effective but their sources of identification (news variables) may have different DGPs (one with more serial correlation than other), which leads to different multipliers.

shock might affect the dynamic responses.

3.3 Differences between DLMs and LPs under persistence

We now consider the two most frequently used methods to estimate impulse responses when a shock is independently identified, DLMs and LPs, and compare the objects that they identify when the shock is persistent. We first consider the case of DLMs. The use of these models is widespread in applied macroeconomics.[11] In the case of system (2), note that we can recover the response function $\mathcal{R}(h)^{DLM}$ using the following regression:[12]

$$y_t = \theta_0 x_t + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \theta_3 x_{t-3} + \theta_4 x_{t-4} + \ldots + e_t, \tag{5}$$

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

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 $h = 0, 1, \ldots, H$:

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

where the sequence of coefficients $\{\delta_{0,h}\}_{h=0}^{H}$ determines the response of the variable of interest $\mathcal{R}(h)^{LP} = \delta_{0,h}$ for each horizon h.[13]

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

[11] See, for example, Romer and Romer (2004), Cerra and Saxena (2008), Romer and Romer (2010), Alesina et al. (2015), Arezki et al. (2017), Coibion et al. (2018) for interesting applications based on DLM methods, or Baek and Lee (2020) for a discussion of their properties. As mentioned in the introduction, these methods are also a special case of more general specifications such as VARs with exogenous variables (or VAR-X). We develop this point further in Section 3.6, when generalizing some of the results of the paper.

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

[13] 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)). See Montiel Olea and Plagborg-Møller (2021) for a recent contribution on inference in LPs.

Proposition 1. *Given the data generating process described by system* (2)*, if the shock x_t is serially uncorrelated, then the response functions identified by DLMs and LPs are equal for all response horizons, that is:*

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

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

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

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

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 last effect (the *persistence* effect of x_t) drives the difference between $\mathcal{R}(h)^{DLM}$ and $\mathcal{R}(h)^{LP}$. In particular, $\mathcal{R}(h)^{LP} = \mathcal{R}(h) = \delta\gamma^h$, while $\mathcal{R}(h)^{DLM} = \mathcal{R}(h)^* = 0$ for all $h \geq 1$.

To understand why LPs, unlike DLMs, incorporate this third effect, consider the LP from equation (6) when $h = 1$ under the special case of $\rho = \delta_1 = 0$ in system (2) (and, therefore, $\rho_h = \delta_{1,h} = 0$ equation (6)) :

$$y_{t+1} = \delta_{0,1} x_t + \xi_{t+1}, \tag{7}$$

where $\delta_{0,1} = \mathcal{R}(1)^{LP}$. 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 (2) to express y_{t+1} as a function of x_t (still maintaining $\rho = \delta_1 = 0$):

$$\begin{aligned} y_{t+1} &= \delta_0 x_{t+1} + u_{t+1} \\ &= \delta_0 \left(\gamma x_t + \varepsilon_{t+1}\right) + u_{t+1} \\ &= \delta_0 \gamma x_t + u^*_{t+1}, \end{aligned}$$

where $u^*_{t+1} = \delta_0 \varepsilon_{t+1} + u_{t+1}$. This shows that the coefficient $\delta_{0,1}$ in equation (7) will also recover the persistence effect of x_t: $\delta_{0,1} = \delta_0 \gamma$. 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 (7), then this effect is absorbed by $\delta_{0,1}$.[14]

When impulse responses are identified using DLMs, the treatment of the persistence of x_t is different. Consider a version of equation (5) expressed in terms of $t+1$:

$$y_{t+1} = \theta_0 x_{t+1} + \theta_1 x_t + \theta_2 x_{t-1} + \theta_3 x_{t-2} + \theta_4 x_{t-3} + \ldots + e_{t+1}. \tag{8}$$

As noted earlier, the sequence of coefficients θ_h determines the response function. Consider the response when $h = 1$, i.e., $\mathcal{R}(1)^{DLM} = \theta_1$. Note that, while we know from system (2) that $\frac{\partial y_{t+1}}{\partial x_t} = \delta_0 \frac{\partial x_{t+1}}{\partial x_t} = \delta_0 \gamma$, the coefficient recovered by θ_1 is indeed $\left.\frac{\partial y_{t+1}}{\partial x_t}\right|_{x_{t+1}} = \delta_0 \left.\frac{\partial x_{t+1}}{\partial x_t}\right|_{x_{t+1}} = 0$. That is, since the DLM controls for x_{t+1}, the persistence effect of x_t is accounted for.

In other words, DLMs identify:

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

while LPs identify:

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

Note that the difference between $\mathcal{R}^{LP}$ and $\mathcal{R}^{DLM}$ is positive (negative) when $\gamma > 0$ ($\gamma < 0$). In empirical applications, γ may be positive or negative.[15]

[14] This omitted variables problem is also briefly mentioned in Alesina et al. (2015) in the particular context of fiscal consolidation plans.

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

3.4 Reestablishing the equivalence between DLMs and LPs

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

3.4.1 Adapting LPs to exclude the effect of serial correlation

As discussed in Section 3.2, a researcher may be interested in recovering responses as if the shock were serially uncorrelated ($\mathcal{R}(h)^*$). However, we have shown that $\mathcal{R}^{LP}(h) \neq \mathcal{R}(h)^*$ 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 (6), or (ii) to replace x_t with the error term that purges out the persistence:

$$\varepsilon_t = x_t - \gamma x_{t-1}. \tag{9}$$

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.3.

A third potential method to exclude the effect of persistence would be recasting system (2) 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 (2021)) the VAR responses would also include an effect due to the persistence of the shock—we explore this in more detail in Section 3.6.

Instead, we propose a method based on the inclusion of leads of the persistent shock variable. In particular, given that the DGP of system (2) poses an AR(1) for x_t, one should regress:

$$y_{t+h} = \rho_h y_{t-1} + \delta_{0,h} x_t + \delta_{1,h} x_{t-1} + \delta_{1^+,h} x_{t+1} + \xi_{t+h}, \tag{10}$$

where $\delta_{0,h}$ is the h-horizon response identified by LPs that include leads of the shock x_t, which we denote as $\mathcal{R}^F(h)$. 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).[16] The most conservative procedure would be to include h leads of the shock in each period h. This is the choice implemented in Section 4, when considering empirical applications.

Proposition 2. *Given the data generating process described by system* (2)*, the response function identified by modified LPs to a shock x_t as described in equation* (10) *is equal to the response as if the shock had no persistence (and to the response obtained from DLMs as in equation* (5)*), that is:*

$\mathcal{R}(h)^F = \mathcal{R}(h)^* = \mathcal{R}(h)^{DLM}$ $\forall$ γ *and* h.

Proof. See Appendix A.2. □

Intuitively, leads of x_t in equation (10) act as controls for the persistence of the shock throughout the response horizon, so that the parameter $\delta_{0,h}$ 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 (2) when $\gamma \neq 0$.

3.4.2 Adapting DLMs to include the effect of persistence

As noted earlier, $\mathcal{R}(h)^{DLM} = \mathcal{R}(h)^*$ regardless of the value of γ. However, in some instances, 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 DLMs to recover these responses. Intuitively, the idea is to compute the impulse responses in system (2) with respect to ε_t instead of x_t.

To see this intuitively, consider a recursive substitution of x_t in system (2) while assuming $\rho = \delta_1 = 0$:

$$y_t = \delta_0 \gamma^t x_0 + \delta_0 \sum_{i=0}^{t} \gamma^i \varepsilon_{t-i} + u_t. \tag{11}$$

The responses of y_t to ε_t, which we denote by $\mathcal{R}(h)^{DLM-per}$, can be obtained from the

[16] See also Lee (2020) for lag order selection in LPs.

coefficients $\tilde{\theta}_h$ in:

$$y_t = \tilde{\theta}_0\varepsilon_t + \tilde{\theta}_1\varepsilon_{t-1} + \tilde{\theta}_2\varepsilon_{t-2} + \tilde{\theta}_3\varepsilon_{t-3} + \tilde{\theta}_4\varepsilon_{t-4} + \ldots + e_t. \tag{12}$$

Proposition 3. *Given the data generating process described by system* (2)*, the response function identified by DLMs of y_t to the innovation ε_t as described in equation* (12) *is equivalent to the response that includes the effects of persistence (and to the response obtained from LPs as in equation* (6)*):*

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

Proof. See Appendix A.3. □

Proposition 3 establishes a direct equivalence between the coefficients obtained from equation (12) and those obtained from LPs in equation (6): $\tilde{\theta}_h = \delta_{0,h}\ \forall\ h$. The former are also related to the coefficients estimated from the DLM in terms of x_t, as in equation (5): $\theta_0 = \tilde{\theta}_0 = \delta$, $\theta_1 = \tilde{\theta}_1 - \gamma\tilde{\theta}_0, \ldots, \theta_h = \tilde{\theta}_h - \gamma\tilde{\theta}_{h-1}$. Intuitively, the response of y_{t+1} to x_t has an overall effect of $\delta_{0,1} = \tilde{\theta}_1$, 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 $\gamma\delta_0$). The *standard* DLM estimation from equation (5), 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.

The Appendix presents simulated examples that complement the above theoretical results. This serves i) to provide intuition and gain further insights (Appendix B.1), ii) show how the results hold when using more empirically-relevant sources of persistence (AppendixB.2), and iii) to show why the inclusions of lags is not sufficient to identify $\mathcal{R}(h)^*$ (Appendix B.3).

3.5 Summary: A guide to practitioners

Here we summarize the lessons from this section, indicating the adjustments required in LPs and DLMs depending on the object of interest. In the presence of a persistent shock,

Table 2: Adapting LPs and DLMs when shocks are persistent

Object of interest / Method	LPs	DLMs
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

the researcher has to determine what object to identify: the response *as if* the shock were uncorrelated ($\mathcal{R}(h)^*$) or the response that includes the effect of persistence ($\mathcal{R}(h)$). As discussed in Section 3.2, deciding for one or the other should depend on what specific question the researcher is trying to address. The researcher also has to decide on what estimation method to use, LPs or DLMs. Table 2 summarizes what adjustments are required in LPs and DLMs depending on the object of interest.

3.6 Generalization to a VAR setting

In this section we explore how the previous results generalize to a multivariate framework. To do so, we follow the existing literature and distinguish two cases. We first consider a VAR where the shock is included in the vector of endogenous variables, and then we turn our attention to a VAR where the shock is included exogenously, as a distributed lag or moving average structure.[17] We show that while both methods yield identical results under the assumption of no persistence in the shock, relaxing this assumption yields results that mirror those obtained in the previous section: a VAR with a shock included as an endogenous variable will identify dynamic effects that include the persistence of the shock as it was the case with LPs[18], while treating the shock as an exogenous variable will yield dynamic effects that exclude the persistence of the shock.

[17] Bloom (2009), Romer and Romer (2010), and Ramey (2011) are examples of studies that include shocks as endogenous variables in a VAR. These specifications are also known as hybrid VARs (see Coibion (2012)). Other papers have opted to include the shock as an exogenous variable in a VAR setting. See for example: Mertens and Ravn (2012) or Favero and Giavazzi (2012))

[18] This also follows from Plagborg-Møller and Wolf (2021), which formally show that VARs (with the shock included in the set of endogenous variables) and LPs identify the same impulse responses.

To illustrate this result, we generalize the original DGP in system (2) by including an additional endogenous variable ($y_{2,t}$), which may have a lagged effect on the main variable of interest ($y_{1,t}$) through parameter A_{21}:

$$\begin{aligned} y_{1,t} &= \rho_1 y_{1,t-1} + A_{21} y_{1,t-2} + \delta_0 x_t + \delta_1 x_{t-1} + \varepsilon_t^{y_1} \\ y_{2,t} &= \rho_2 y_{2,t-1} + C_0 x_t + C_1 x_{t-1} + \varepsilon_t^{y_2} \\ x_t &= \gamma x_{t-1} + \varepsilon_t^{x}. \end{aligned} \tag{13}$$

First, we explore the case where the shock x_t is included with the rest of the endogenous variables in the VAR. Note that in that case the system (13) 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:

$$\begin{bmatrix} 1 & 0 & 0 \\ -\delta_0 & 1 & 0 \\ -C_0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \gamma & 0 & 0 \\ \delta_1 & \rho_1 & A_{21} \\ C_1 & 0 & \rho_2 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{1,t-1} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{x} \\ \varepsilon_t^{y_1} \\ \varepsilon_t^{y_2} \end{bmatrix}. \tag{14}$$

An econometrician would estimate the following reduced-form VAR:

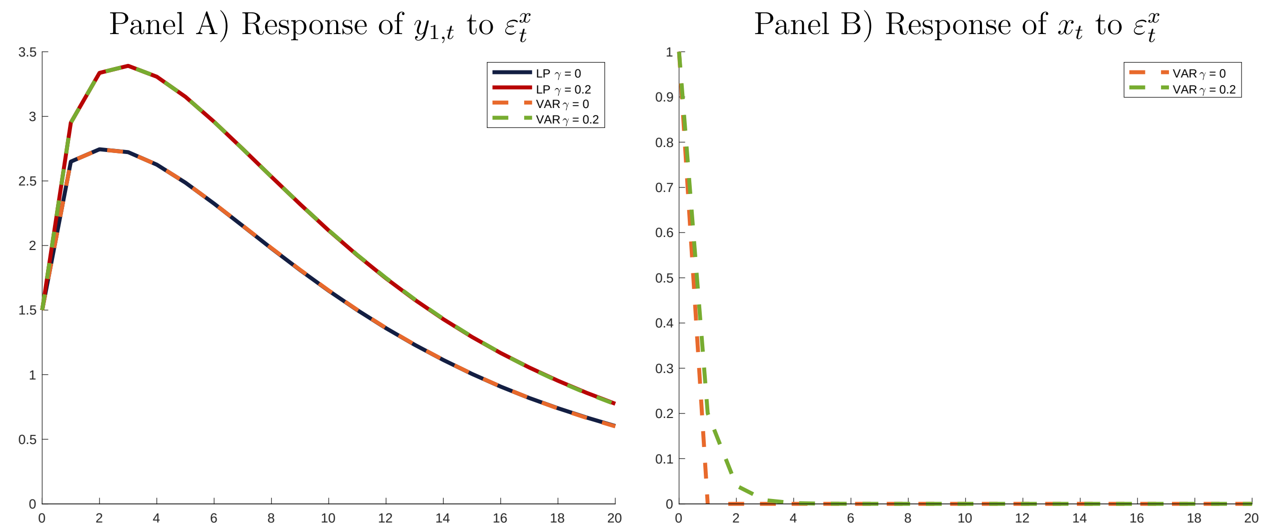
$$\boldsymbol{Y}_t = \boldsymbol{B} \boldsymbol{Y}_{t-1} + \boldsymbol{u}_t, \tag{15}$$

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 DGP given by equation (13) 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 $\boldsymbol{u}_t$.

Panel A of Figure 1 shows the dynamic response of $y_{1,t}$ to an innovation in to ε_t^{x}, when assuming two different cases for the persistence of the shock: $\gamma = 0$ (dashed orange line) and $\gamma > 0$ (dashed green line).[19] Note that for comparison, we also plot the response estimated

[19]We consider a calibration of system (13) with $\rho^1 = 0.9$, $\rho^2 = 0.7$, $\delta_0 = 1.5$, $\delta_1 = 1$, $C_0 = 1$, $C_1 = 0.5$,

Figure 1: Responses in a VAR 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 (15), 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).

by LPs of $y_{1,t}$ to a contemporaneous shock in x_t considering the same two cases of persistence of x_t (solid blue and red lines, respectively).

The results show that when $\gamma > 0$, the variable x_t accumulates over time the initial innovation in ε_t^x (see Panel B in Figure 1). In turn, this effect propagates to $y_{1,t}$ through i) the lagged effect of x_t on $y_{1,t}$, ii) the effect that x_t has through the persistence of $y_{1,t}$, and iii) the effect that $y_{2,t}$ has on $y_{1,t}$. As it becomes explicit from system (14), the persistence of the shock x_t is endogenously estimated in the VAR and incorporated in the the responses of the endogenous variables. Hence, the dynamic responses of $y_{1,t}$ becomes higher when there is positive persistence in x_t. Note that Panel A in Figure 1 also shows that the responses obtained from a VAR with a shock treated as an endogenous variable mimic the results shown earlier in the section when considering the use of LPs.[20]

We now turn our attention to an alternative method to include the shock in the VAR. We

$A_2 1 = 0.3$ and either $\gamma = 0$ or $\gamma = 0.2$.

[20]This is formally shown in Plagborg-Møller and Wolf (2021), which state that LPs and VARs (when the shock is treated endogenously) identify the same impulse-responses.

start again from system (13), but following Mertens and Ravn (2012) or Favero and Giavazzi (2012), we instead consider x_t as an exogenous element in the VAR (which is sometimes referred to as VAR-X):

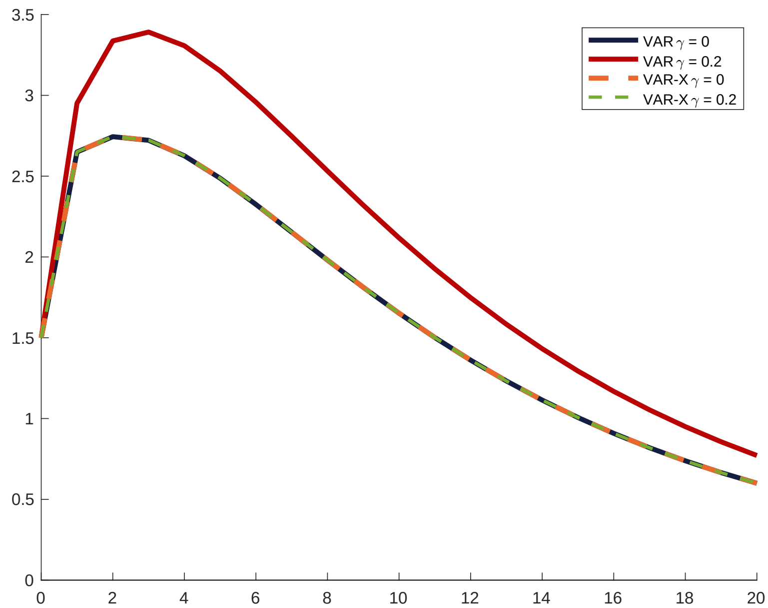
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_1 & A_{21} \\ 0 & \rho_2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} \delta_0 \\ C_0 \end{bmatrix} x_t + \begin{bmatrix} \delta_1 \\ C_1 \end{bmatrix} x_{t-1} + \begin{bmatrix} \varepsilon_t^{y_1} \\ \varepsilon_t^{y_2} \end{bmatrix}. \tag{16}$$

The results of this estimation are presented in Figure 2. Note that when there is no persistence in the variable x_t, the response of $y_{1,t}$ is the same regardless of including the shock as an exogenous variable (dashed orange line) or, as we did before, as an endogenous one (solid blue line). However, this equivalence does not hold when $\gamma \neq 0$, since the VAR with the shock as an exogenous variable continues to recover the same response as if $\gamma = 0$ (dashed green line) while, as shown before, the VAR with the shock as an endogenous variable includes the effect of the persistence of x_t on the response of $y_{1,t}$. To explain the difference, note that when constructing impulse responses in this framework, we are shocking x_t (not $\varepsilon^{y_{1,t}}$), and therefore we are constructing a different thought experiment than in the previous case.

These simulations illustrate that the results shown in Proposition 1 generalize to a multivariate setting. Hence, the dynamic responses of a variable to a persistent shock will depend on how this shock is included in the VAR.

Finally, it is worth noting that the representation of the DGP as a multivariate process highlights the role that other variables can have in transmitting the effect of a shock. Hence, parallel to what was described earlier in this section, introducing leads of $y_{2,t}$ yields counterfactual responses $\mathcal{R}(h)^*$ of $y_{1,t}$, where the path of $y_{2,t}$ is held constant throughout the response horizon, that is, the channel of propagation of the shock x_t through $y_{2,t}$ is shut down. The construction of this counterfactual is similar in spirit to that of Cloyne et al. (2020) (where the authors use a Blinder-Oaxaca approach) and opens the possibility for an heuristic exploration of the transmission mechanisms of shocks. This avenue could be par-

Figure 2: Responses when the shock is included as an endogenous or exogenous variable in a VAR



The figure shows the VAR responses of $y_{1,t}$ to ε_t^x estimated from (15) 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 (as in system (14)) and dashed line shows the responses when the same shock is included as an exogenous variable in the VAR (as in system (16)) .

ticularly informative to understand which economic models can bring the dynamic responses closer to the data. We explore its empirical relevance in the context of government spending shocks at the end of Section 4.

3.7 Local projections with instrumental variables

There is increasing attention to the use of external sources of variation as instruments in LPs.[21] Intuitively, these variables can be included either as direct measures of shocks or as instruments depending on the researcher's judgment (see Gertler and Karadi (2015) and Ottonello and Winberry (2020) for contrasting examples). When they are considered to be instruments in local projections (LP-IV), the persistence in these variables may lead to the violation of the lead/lag exogeneity condition (Stock and Watson (2018)), affecting the identification of the object of interest.[22]

In a LP-IV setting it becomes relevant to understand what causes the serial correlation in the instrument. While a full characterization lies beyond the scope of this paper, we briefly discuss two cases to guide researchers in dealing with persistence in a instrument when using LPs. We focus on the identification of $\mathcal{R}(h)^*$, which is more likely to be the object of interest in a LP-IV context.

First, if the source of persistence in the instrument is independent of the past and future realizations of the shock in the system, then the lead/lag exogeneity assumption is not violated. This is the general case laid out by Stock and Watson (2018). In this regard, using narrative measures as instruments might not problematic when the researcher is interested in $\mathcal{R}(h)^*$.

Second, if the instrument inherits its persistence from the shock (that is, the underlying

[21] See Ramey and Zubairy (2018) for an example. Related to this, Ramey (2016) discusses the distinction between shock, innovation, and instrument. Barnichon and Mesters (2020) show how independently identified shocks can be used as instruments to estimate the coefficients of structural forward looking macroeconomic equations.

[22] Stock and Watson (2018) state that 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. Also, the instrument should not be correlated with any lead or lag of any of the shocks in the system (the lead/lag exogeneity restriction).

shock is persistent and, in turn, this reflects on the instrument) the lead/lag exogeneity can only be re-established by including leads of the shock. The intuition follows from the results in the previous sections: when the instrument takes its persistence form the shock, the inclusion of lags will not be sufficient to account for the intermediate behaviour of the shock between time t and $t + h$. In this case, only the inclusion of leads would bring back lead/lag exogeneity and allow identification of $\mathcal{R}(h)^*$.

In sum, the implications of persistence in LP-IV settings are nuanced as they depend on the source of persistence in the model. As mentioned in the above examples, sometimes lags may be sufficient to identify $\mathcal{R}(h)^*$ while other DGPs may imply that leads are needed. Following from the previous results in the section, our insight is that when the researcher is unsure of the nature of the serial correlation in the instrument, the inclusion of leads and the computation of $\mathcal{R}(h) - \mathcal{R}(h)^*$ might help to detect to which extent the dynamic responses depend on the serial correlation of the instrument. This recommendation is further reassured by the conclusion obtained when investigating actual empirical applications where the shock is not persistent. In this case, the *unnecessary* inclusion of leads to recover $\mathcal{R}(h)^*$ does not seem to carry a penalty in terms of inference (see Appendix C.2).

4 Application

In this section we use the empirical work of Ramey and Zubairy (2018) to show the quantitative relevance of serial correlation in an actual example. We do so by computing two types of IRFs, $\mathcal{R}(h)$ and $\mathcal{R}(h)^*$, as described above. As discussed before, the difference between both estimates, $\mathcal{R}(h) - \mathcal{R}(h)^*$, is the part of the dynamic response which is due to the effects of persistence in the shock.[23] In Appendix C we consider additional applications based on Guajardo et al. (2014) and Romer and Romer (2010).

[23] At the end of this section we show the results of an alternative decomposition of $\mathcal{R}(h)$, where we highlight the part of the dynamic response which is attributable to the effect of a third variable.

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.[24] This series, plotted in panel D of Figure D2, has a positive autocorrelation of 18.4% (47.0% in the subsample after WWII).

Ramey and Zubairy (2018) use LPs to estimate the response of output and government spending to a shock in future defense spending.[25] We follow their same approach and sample and estimate the following equations for output (y_t) and government spending (g_t):

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

where z_t includes P lags of y_t, g_t and $shock_t$. Note that, following the discussion in previous sections, we include h leads of the variable $shock_t$. In particular, for each horizon h we include h leads.

To replicate Ramey and Zubairy (2018)'s estimates, we set $\gamma_{f,h}^y = \gamma_{f,h}^g = 0, \forall\ f, h$. The black, solid line in Figure 3 represents the estimated responses of output (left panel) and government spending (right panel) to the shock. As noted in Section 3, these dynamic responses are the equivalent to the $\mathcal{R}(h)$ as defined in equation (3) (with the only difference being that Ω_{t-1} includes now the past history of z_t). The results closely resemble those in Ramey and Zubairy (2018) (Figure 5 of their paper).[26]

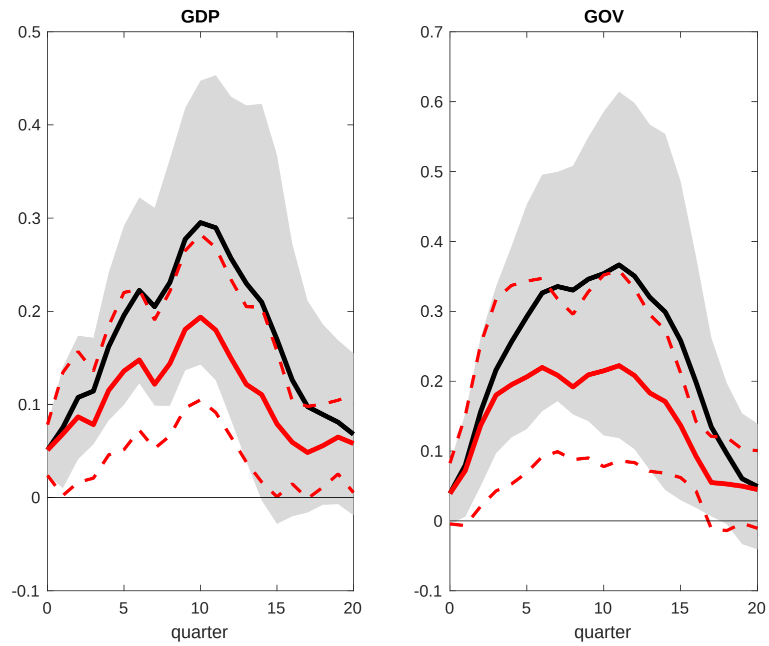
Next, we allow $\gamma_{f,h}^y \neq 0$ and $\gamma_{f,h}^g \neq 0$. As discussed in Section 3, this amounts to estimating $\mathcal{R}(h)^*$ as defined in equation (4). In the red lines in Figure 3, we observe that

[24] 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).

[25] See Gonçalves et al. (2021) for an alternative LP estimator in non-linear settings.

[26] 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).

Figure 3: Output and government spending responses, with and without leads



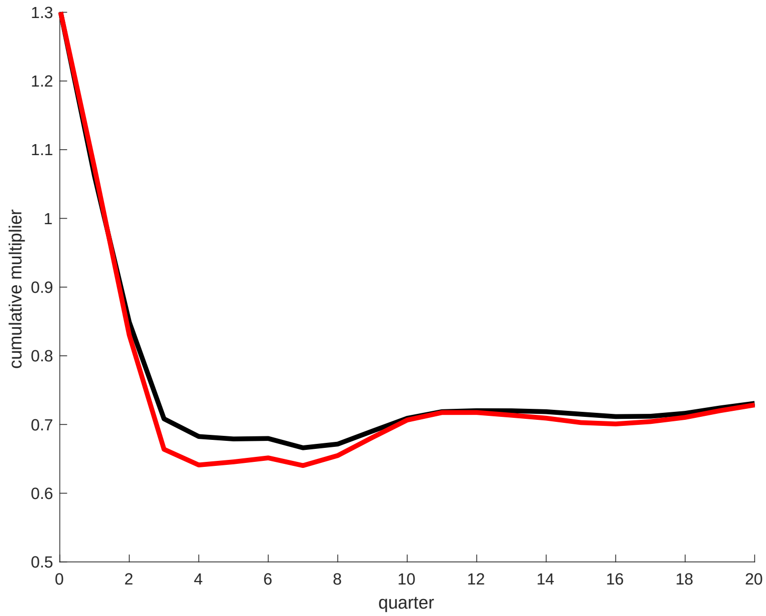
Black lines show the results of estimating the system (17) without including any lead (as in Ramey and Zubairy (2018)). Grey areas represent 68 and 95% Newey-West confidence intervals for these estimates. Red solid lines represent the results of estimations when including h leads of the Ramey and Zubairy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

the dynamic 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. The large observed difference between $\mathcal{R}(h)$ and $\mathcal{R}(h)^*$ suggests that the persistence of the news variable plays a non-negligible role in explaining the dynamic transmission of the fiscal shock to output and government spending.

Whether to include leads or not also has implications for inference. The 95% confidence intervals when leads are included (shown in dashed lines in Figure 3) are substantially narrower than when they are not (grey areas in Figure 3). 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

Figure 4: 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.

and Zubairy (2018) use the cumulative multiplier, computed as:[27]

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

We find that this statistic is not substantially affected by persistence of the shock (Figure 4). 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.

Note that, even though the multiplier does not change much when accounting for persistence, the fact that the expected responses of output and government spending do change substantially is very relevant from a policy-maker point of view. For example, a higher response of government spending can affect other important variables such as public debt or

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

future changes in tax liabilities. Also note that it is just a feature of this concrete example that the effects of persistence on the output and government spending cancel out: in terms of equation (17), for the cumulative multiplier to cancel the effects of persistence, $\gamma^y_{f,h}$ and $\gamma^g_{f,h}$ must have the same implications for the response of output and government spending, respectively. In fact, the next paragraphs show an example when this is not the case—the cumulative multiplier does change substantially in the non-linear case.

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 different in expansions and recessions.[28] For this, we follow Ramey and Zubairy (2018) and estimate a series of non-linear LPs:

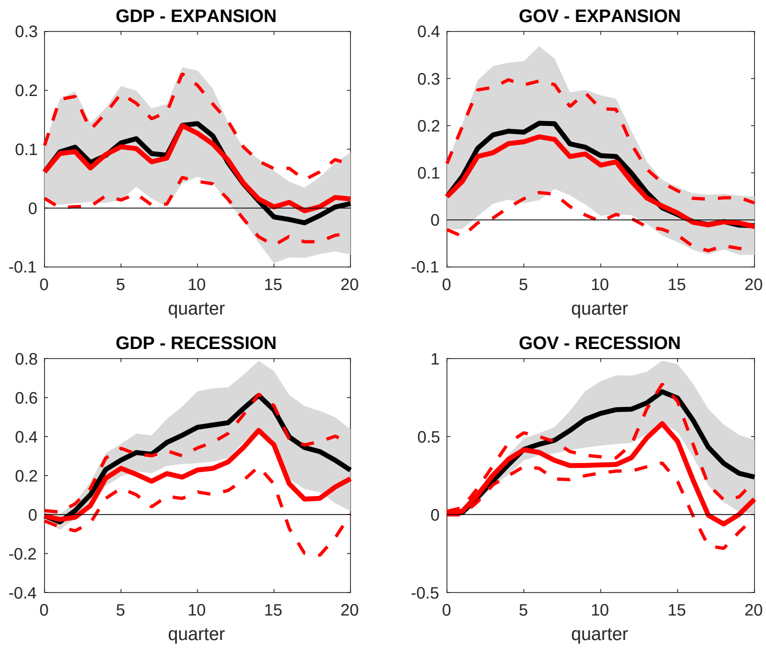
$$
\begin{aligned}
x_{t+h} = S_{t-1} &\left[\alpha_{A,h} + \sum_{j=1}^{P} \rho_{A,j,h} z_{t-j} + \beta_{A,h} shock_t + \sum_{f=1}^{h} \delta_{A,f,h} shock_{t+f}\right] + \\
(1 - S_{t-1}) &\left[\alpha_{B,h} + \sum_{j=1}^{P} \rho_{A,j,h} z_{t-j} + \beta_{B,h} shock_t + \sum_{f=1}^{h} \delta_{B,f,h} shock_{t+f}\right]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

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 (and 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 $\delta_{A,f,h} = \delta_{B,f,h} = 0 \ \forall f, h$, which identifies $\mathcal{R}(h)$. Our results, shown in

[28] 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 (2022) 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.

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



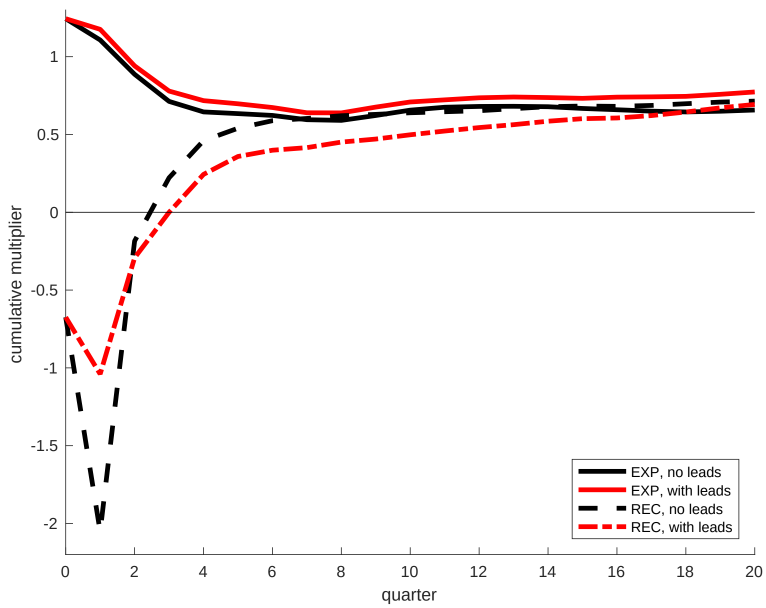
Black lines show the results from system of equations (19) without including any lead (as in Ramey and Zubairy (2018)). Grey areas represent 68 and 95% Newey-West confidence intervals for these estimates. Red solid lines represent the results of estimations when including h leads of the Ramey's news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

Figure 5 in black lines, resemble very closely those from the authors. Next, we repeat the experiment accounting for potential persistence, that is, including leads of the shock (identifying $\mathcal{R}(h)^*$). 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, if persistence is taken into account, responses 2-3 years after recessions are half in size than if persistence were not taken into account. Or, in other words, persistence in the shock is responsible for up to 50% of the dynamic transmission of the shock during recessions.

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, $\mathcal{R}(h)^*$) or not (black solid line, $\mathcal{R}(h)$). 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 excluded from the dynamic responses (red dashed line), the cumulative multiplier is -1 (instead of -2) and becomes positive after the first year. Furthermore, the multiplier during recessions remains lower than the multiplier during expansions for a much longer period. When the persistence is not accounted for, 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 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, i.e. the state is always in a recession or in an expansion). 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 D3. 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

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.

in conjunction with LPs can help understand interesting counterfactual exercises and shed light on the dynamic transmission of shocks.

Exploring the transmission of the shock Next, we empirically test the mechanism mentioned at the end of Section 3.6. In particular, we are interested in exploring how much of the (linear) response of output to a government spending shock is due to effects through the labor market.[29] To do so, we first re-estimate the effects of government spending as described in equation (17) but including the unemployment rate as an additional control variable. The resulting cumulative multiplier, which is roughly the same size of the linear estimates in Ramey and Zubairy (2018), is shown in black in Figure 7.[30]

To explore how much of the transmission of the shock is lost when the unemployment rate is not allowed to move, we include h leads of this variable in each of the LPs. This means that while we allow for the unemployment rate to respond on impact, this variable is kept constant throughout the rest of the response horizon. The cumulative multiplier, shown in red in Figure 7 is greatly reduced: after two years, the multiplier is almost three times lower when shutting down the transmission through the labor market.

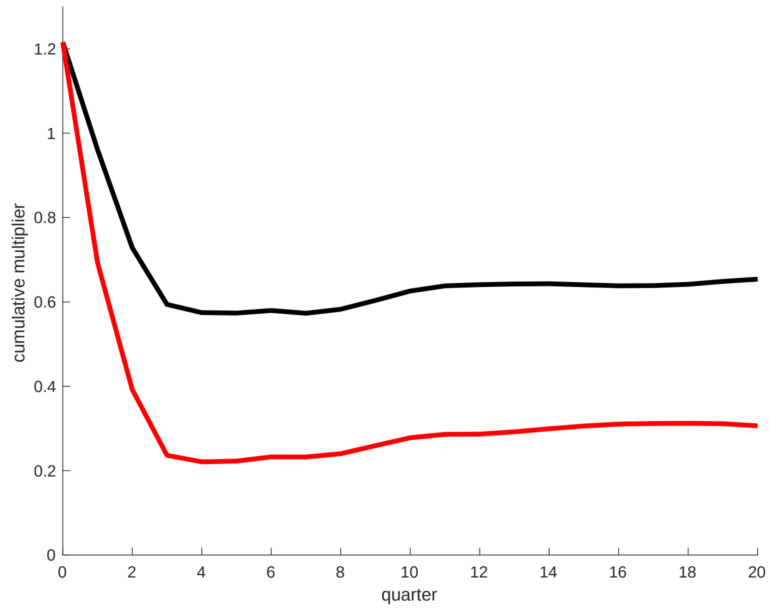
5 Conclusions

We have shown that the presence of some degree of persistence in shocks leads to the estimation of different responses when using LPs *versus* traditional methods based on DLMs. For a researcher interested in the response *as if* the shock were not persistent, DLMs yield the desired object, but LPs need to be adapted. The opposite is true if the object of interest is the response to the shock "as it is". Regardless of which is the thought experiment that the researcher seeks to carry out, the difference between both types of responses is informative about how much of the dynamic transmission of a shock is due to the presence of persistence.

[29] Both the intensive and extensive margin of labor are key channels of transmission of fiscal policies. See, for example, Baxter and King (1993).

[30] Appendix Figure D4 shows the response of all variables.

Figure 7: Government spending multiplier including leads of the unemployment rate



The black line shows the government spending cumulative multiplier (as in Ramey and Zubairy (2018)). The red line shows the cumulative multiplier when including leads of the unemployment rate.

As shown in the previous section, the use of leads can be generalized to other interesting contexts, as it allows to shut down channels of transmission. For example, one may be interested in the effects of monetary policy shocks on output due to a particular instrument while holding other variables (e.g., changes to fiscal policy) constant. 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. The inclusion of leads achieves a similar goal in LPs, hence allowing to construct interesting macroeconomic experiments. We leave these questions for future research.

References

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

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

Alloza, M. (2022). Is fiscal policy more effective during recessions? *International Economic Review*, 63(3):1271–1292.

Alloza, M. and Sanz, C. (2021). Jobs multipliers: Evidence from a large fiscal stimulus in Spain. *Scandinavian Journal of Economics*, 123(3):751–779.

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. (2020). A Guide to Single Equation Regressions for Impulse Response Estimations. *Working Paper, University of California, Berkeley.*

Barnichon, R., Debortoli, D., and Matthes, C. (2022). Understanding the size of the government spending multiplier: It's in the sign. *Review of Economic Studies*, 89(1):87–117.

Barnichon, R. and Mesters, G. (2020). Identifying modern macro equations with old shocks. *Quarterly Journal of Economics*, 135(4):2255–2298.

Baxter, M. and King, R. G. (1993). Fiscal policy in general equilibrium. *American Economic Review*, pages 315–334.

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.

Cerra, V. and Saxena, S. C. (2008). Growth Dynamics: The Myth of Economic Recovery. *American Economic Review*, 98(1):439–57.

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

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.

Cloyne, J. S., Jordà, Ò., and Taylor, A. M. (2020). Decomposing the Fiscal Multiplier. *NBER Working Paper*, 26939.

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

Coibion, O., Gorodnichenko, Y., and Ulate, M. (2018). The Cyclical Sensitivity in Estimates of Potential Output. *Brookings Papers on Economic Activity*, page 343.

Devries, P., Guajardo, J., Leigh, D., and Pescatori (2011). A New Action-Based Dataset of Fiscal Consolidation. *IMF Working Paper*, 11/128.

Dube, A., Girardi, D., Jorda, O., and Taylor, A. M. (2023). A local projections approach to difference-in-differences event studies. *NBER WP 31184*.

Dufour, J.-M. and Renault, E. (1998). Short Run and Long Run Causality in Time Series: Theory. *Econometrica*, 66(5):1099–1125.

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

Favero, C. and Giavazzi, F. (2012). Measuring Tax Multipliers: The Narrative Method in Fiscal VARs. *American Economic Journal: Economic Policy*, 4(2):69–94.

Fisher, J. D. and Peters, R. (2010). Using Stock Returns to Identify Government Spending Shocks. *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.

Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2021). Impulse response analysis for structural dynamic models with nonlinear regressors. *Journal of Econometrics*, 225(1):107–130.

Gorodnichenko, Y. and Lee, B. (2020). Forecast error variance decompositions with local projections. *Journal of Business & Economic Statistics*, 38(4):921–933.

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

Granger, C. W. and Machina, M. J. (2006). Chapter 2 Forecasting and Decision Theory. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 81 – 98. Elsevier.

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

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.

Lee, B. (2020). Lag Order Selection in Local Projections. *Working Paper, Hong Kong University of Science and Technology.*

Lee, T.-H. (2008). Loss functions in time series forecasting. *International Encyclopedia of the Social Sciences*, pages 495–502.

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.

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

Montiel Olea, J. L. and Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, 89(4):1789–1823.

Ottonello, P. and Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502.

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. (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, 89(2):955–980.

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

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. (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.

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

First, we derive the impulse-response functions identified by LP. Consider the DGP (2) (rewritten here for convenience):

$$\begin{aligned} y_t &= \rho y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + u_t \\ x_t &= \gamma x_{t-1} + \varepsilon_t. \end{aligned} \tag{A.1}$$

Consider the LP:

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

where $\delta_{0,h} = \mathcal{R}(h)^{LP}$ is the linear projection, that is:

$$\delta_{0,h} = \frac{Cov(y_{t+h}, x_t | y_{t-1}, x_{t-1})}{Var(x_t | y_{t-1}, x_{t-1})}.$$

Note that iterating in the DGP A.1, y_{t+h} can be expressed as a function of x_t and the m.d.s. ε_t and u_t:

$$\begin{aligned} y_{t+h} = [&\delta_0 \gamma^h + (\delta_0 \rho + \delta_1)\gamma^{h-1} + (\delta_0 \rho^2 + \delta_1 \rho)\gamma^{h-2} + \\ &\ldots + (\delta_0 \rho^{h-1} + \delta_1 \rho^{h-2})\gamma + (\delta_0 \rho^h + \delta_1 \rho^{h-1})] x_t + \\ &\ldots + f(u_t, \ldots, u_{t+h}) + g(\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}). \end{aligned}$$

where $f(u_t, \ldots, u_{t+h})$ is a linear function of $u_t, \ldots, u_{t+h}$, and $g(\varepsilon_{t+1}, \ldots, \varepsilon_{t+h})$ is a linear function of $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$.

Or, in a more compact form,

$$y_{t+h} = \sum_{j=0}^{h} \alpha_j \gamma^{h-j} x_t + f(u_t, \ldots, u_{t+h}) + g(\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}),$$

where

$$\alpha_j = \delta_0 \quad \text{if} \quad j = 0$$
$$\alpha_j = \delta_0 \rho^j + \delta_1 \rho^{j-1} \quad \text{if} \quad j > 0.$$

Plugging this into the LP estimator:

$$\begin{aligned} \delta_{0,h} &= \frac{Cov(\sum_{j=0}^{h} \alpha_j \gamma^{h-j} x_t + f(u_t, \ldots, u_{t+h}) + g(\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}), x_t)|y_{t-1}, x_{t-1})}{Var(x_t|y_{t-1}, x_{t-1})} \\ &= \sum_{j=0}^{h} \alpha_j \gamma^{h-j} \frac{Cov(x_t, x_t|.)}{Var(x_t|.)} + \frac{Cov(f(u_t, \ldots, u_{t+h}), x_t|.)}{Var(x_t|.)} + \frac{Cov(g(\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}), x_t|.)}{Var(x_t|.)}, \end{aligned} \tag{A.2}$$

where the last two terms are equal to 0 because the errors are m.d.s. and independent of each other.

Hence,

$$\delta_{0,h} = \sum_{j=0}^{h} \alpha_j \gamma^{h-j} = \mathcal{R}(h).$$

As a second step, we derive the dynamic responses identified by DLMs. Consider:

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

Forwarding this equation:

$$y_{t+h} = \theta_0 x_{t+h} + \theta_1 x_{t+h-1} + \theta_2 x_{t+h-2} + \ldots + \theta_h x_t + \ldots + u_{t+h}.$$

We are interested in $\frac{\partial y_{t+h}}{\partial x_t} | x_{t+h}, \ldots, x_{t+1}, \ldots, x_{t-1} = \theta_h = \mathcal{R}(h)^{DLM}$.

Hence,

$$\theta_h = \frac{Cov(y_{t+h}, x_t | x_{t+h}, \ldots, x_{t+1}, \ldots, x_{t-1}, \ldots)}{Var(x_t | x_{t+h}, \ldots, x_{t+1}, \ldots, x_{t-1}, \ldots)}. \tag{A.3}$$

Now, note that the DGP given by equation (A.1) can be converted into a moving average representation in terms of the shocks x such that, at time $t + h$:

$$y_{t+h} = \alpha_0 x_{t+h} + \alpha_1 x_{t+h-1} + \ldots + \alpha_h x_t + \alpha_{h+1} x_{t-1} + \ldots + v_{t+h},$$

where

$$v_{t+h} = u_{t+h} + \rho u_{t+h-1} + \rho^2 u_{t+h-2} + \ldots$$

Plugging this into (A.3):

$$\begin{aligned} \theta_h \quad &= \quad \frac{Cov(\alpha_0 x_{t+h} + \alpha_1 x_{t+h-1} + \ldots + \alpha_{h-1} x_{t+1}, x_t | .)}{Var(x_t | .)} + \frac{Cov(\alpha_h x_t, x_t | .)}{Var(x_t | .)} \\ &+ \quad \frac{Cov(\alpha_{h+1} x_{t-1} + \alpha_{h+2} x_{t-2} + \ldots, x_t | .)}{Var(x_t | .)} + \frac{Cov(v_{t+h}, x_t | .)}{Var(x_t | .),} \end{aligned}$$

where the first term equals zero because we condition on $x_{t+1}, \ldots, x_{t+h}$, the third term equals zero because we condition on $x_{t-1}, x_{t-2}, \ldots$, and the final term equals zero because v_t is a linear combination of u_t, an m.d.s.

Hence,

$$\begin{aligned}\theta_h &= \alpha_h \frac{Cov(x_t, x_t|.)}{Var(x_t|.)} = \delta_0 = \mathcal{R}(h)^* \quad \text{if } h = 0 \\ \text{and } \theta_h &= \delta_0 \rho^h + \delta_1 \rho^{h-1} = \mathcal{R}(h)^* \quad \text{if } h > 0.\end{aligned}$$

■

A.2 Proof of Proposition 2

Consider the LP:

$$y_{t+h} = \rho_h y_{t-1} + \delta_{0,h} x_t + \delta_{1,h} x_{t-1} + \delta_{1+,h} x_{t+1} + \xi_{t+h}.$$

We are interested in the coefficient $\delta_{0,h}$:

$$\delta_{0,h} = \frac{Cov(y_{t+h}, x_t | y_{t-1}, x_{t+1}, x_{t-1})}{Var(x_t | y_{t-1}, x_{t+1}, x_{t-1})}.$$

Note that, from the DGP given by equation (A.1), y_{t+h} can be written as a function of x_{t+1}, x_t, past values of x_t, ε_t and u_t:

$$\begin{aligned}y_{t+h} &= \alpha_0 \gamma^{h-1} + \sum_{j=1}^{h-1} \alpha_j \gamma^{h-j} x_{t+1} + \\ &\quad \alpha_h x_t + \sum_{j=1}^{t+1} \alpha_j x_{t-j} + \bar{f}(u_0, \ldots, u_{t+h}) + \bar{g}(\varepsilon_{t+2}, \ldots, \varepsilon_{t+h}),\end{aligned}$$

where

$$\begin{aligned}\alpha_0 &= \delta_0 \qquad \text{(A.4)} \\ \alpha_j &= \delta_0 \rho^j + \delta_1 \rho^{j-1} \text{ if } j > 0,\end{aligned}$$

$\bar{f}(u_0, \ldots, u_{t+h})$ is a linear function of $u_0, \ldots, u_{t+h}$, and $\bar{g}(\varepsilon_{t+2}, \ldots, \varepsilon_{t+h})$ is a linear function of $\varepsilon_{t+2}, \ldots, \varepsilon_{t+h}$.

Note that, from the DGP, we have:

$$x_{t+h} = \gamma^h x_t + \sum_{j=0}^{h-1} \gamma_j \varepsilon_{t+h-j},$$

so we can substitute for the past values of x_t in the previous equation:

$$\begin{aligned} y_{t+h} &= \alpha_0 \gamma^{h-1} + \sum_{j=1}^{h-1} \alpha_j \gamma^{h-j} x_{t+1} \\ &\quad + \alpha_h x_t + \gamma^t x_0 + \sum_{j=0}^{t-1} \gamma^j \varepsilon_{t-j} + \bar{f}(u_0, \ldots, u_{t+h}) + \bar{g}(\varepsilon_{t+2}, \ldots, \varepsilon_{t+h}) \end{aligned}$$

so that

$$\begin{aligned} y_{t+h} &= \alpha_0 \gamma^{h-1} + \sum_{j=1}^{h-1} \alpha_j \gamma^{h-j} x_{t+1} \\ &\quad + \alpha_h x_t + \gamma^t x_0 + \bar{f}(u_0, \ldots, u_{t+h}) + \bar{\bar{g}}(\varepsilon_0, \ldots, \varepsilon_{t+h}), \end{aligned}$$

where $\bar{\bar{g}}(\varepsilon_0, \ldots, \varepsilon_{t+h})$ is a linear function of $\varepsilon_0, \ldots, \varepsilon_{t+h}$.

Plugging this back into the LP estimator:

$$\begin{aligned} \delta_{0,h} = \frac{Cov(\alpha_0 \gamma^{h-1}, x_t|.)}{Var(x_t|.)} &+ \frac{Cov(\sum_{j=1}^{h-1} \alpha_j \gamma^{h-j} x_{t+1}, x_t|.)}{Var(x_t|.)} \\ &+ \alpha_h \frac{Cov(x_t, x_t|.)}{Var(x_t|.)} + \frac{Cov(\gamma x_0, x_t|.)}{Var(x_t|.)} \\ + \frac{Cov(\bar{f}(u_0, \ldots, u_{t+h}), x_t|.)}{Var(x_t|.)} &+ \frac{Cov(\bar{\bar{g}}(\varepsilon_0, \ldots, \varepsilon_{t+h}), x_t|.)}{Var(x_t|.)}, \end{aligned}$$

where the first and fourth terms equal zero because they include a constant, the second term equals zero because we control for x_{t+1}, and the last two terms equal zero because the errors are a m.d.s. and independent from each other

Hence,

$$\delta_{0,h} = \mathcal{R}(h)^F = \alpha_h = \delta_0\rho^h + \delta_1\rho^{h-1} = \mathcal{R}(h)^{DLM} = \mathcal{R}(h)^*,$$

where the last equality was shown in Proof A.1. ■

A.3 Proof of Proposition 3

Consider the DLM regression:

$$y_t = \tilde{\theta}_0\varepsilon_t + \tilde{\theta}_1\varepsilon_{t-1} + \tilde{\theta}_2\varepsilon_{t-2} + \ldots + \eta_t,$$

in $t+h$:

$$y_{t+h} = \tilde{\theta}_0\varepsilon_{t+h} + \tilde{\theta}_1\varepsilon_{t+h-1} + \ldots + \tilde{\theta}_h\varepsilon_t + \tilde{\theta}_{h+1}\varepsilon_{t-1} + \ldots + \eta_t.$$

We are interested in $\tilde{\theta}_h$:

$$\tilde{\theta}_h = \frac{Cov(y_{t+h}, \varepsilon_t | \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{t-1}, \ldots)}{Var(\varepsilon_t | \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{t-1}, \ldots)}.$$

Since ε_t is a m.d.s., the conditioning set is redundant:

$$\tilde{\theta}_h = \frac{Cov(y_{t+h}, \varepsilon_t)}{Var(\varepsilon_t)}.$$

Using expression (A.2),

$$\tilde{\theta}_h = \frac{Cov(\sum_{j=0}^{h} \alpha_j \gamma^{h-j} x_t + \bar{\bar{f}}(u_t, \dots, u_{t+h}) + \bar{\bar{g}}(\varepsilon_{t+1}, \dots, \varepsilon_{t+h}), \varepsilon_t)}{Var(\varepsilon_t)},$$

where

$$\begin{aligned} \alpha_0 &= \delta_0 \\ \alpha_j &= \delta_0 \rho^j + \delta_1 \rho^{j-1} \text{ if } j > 0, \end{aligned}$$

$\bar{\bar{f}}(u_t, \dots, u_{t+h})$ is a linear function of $u_t, \dots, u_{t+h}$, and $\bar{\bar{g}}(\varepsilon_{t+1}, \dots, \varepsilon_{t+h})$ is a function of $\varepsilon_{t+1}, \dots, \varepsilon_{t+h}$.

Since the error terms are m.d.s. and independent from each other:

$$Cov(\bar{\bar{f}}(u_t, \dots, u_{t+H}), \varepsilon_t) = Cov(\bar{\bar{g}}(\varepsilon_{t+1}, \dots, \varepsilon_{t+H}), \varepsilon_t) = 0.$$

Hence,

$$\tilde{\theta}_h = \sum_{j=0}^{h} \alpha_j \gamma^{h-j} \frac{Cov(x_t, \varepsilon_t)}{Var(\varepsilon_t)},$$

Note that since, as mentioned before,

$$x_{t+h} = \gamma^h x_t + \sum_{j=0}^{h-1} \gamma^j \varepsilon_{t+h-j},$$

then

$$x_t = \gamma^t x_0 + \sum_{j=0}^{t} \gamma^j \varepsilon_{t-j}.$$

Hence,

$$\begin{aligned}
\tilde{\theta}_h &= \sum_{j=0}^{h} \alpha_j \gamma^{h-j} \frac{\gamma^h Cov(x_0, \varepsilon_t) + Cov(\varepsilon_t + \gamma\varepsilon_{t-1} + \gamma^2\varepsilon_{t-2} + \ldots, \varepsilon_t)}{Var(\varepsilon_t)} = \\
&= \sum_{j=0}^{h} \alpha_j \gamma^{h-j} \frac{Cov(\varepsilon_t, \varepsilon_t)}{Var(\varepsilon_t)} = \sum_{j=0}^{h} \alpha_j \gamma^{h-j} = \mathcal{R}(h)^{DLM-per}.
\end{aligned}$$

This, combined with the result from Proposition 1, yields:

$$\mathcal{R}(h)^{DLM-per} = \mathcal{R}(h)^{LP} = \mathcal{R}(h)$$

■

B Examples and additional results

B.1 Main examples

In this subsection, we perform stochastic simulations of the asymptotic behavior of the impulse response functions using both LPs and DLMs. We simulate system (2) for 100 million periods, setting $\delta_0 = 1.5$, $\delta_1 = 1$ and $\rho = 0.9$. We recover the dynamic responses of y_t to the shock x_t using LPs as in equation (10).

We consider three cases: (i) no persistence ($\gamma = 0$), without including leads in the estimation (i.e., setting $\delta_{1+,h} = 0$); (ii) some persistence ($\gamma = 0.2$) and still $\delta_{1+,h} = 0$; (iii) some persistence ($\gamma = 0.2$) and including a lead of the explanatory variable (i.e., allowing $\delta_{1+,h} \neq 0$).[1]

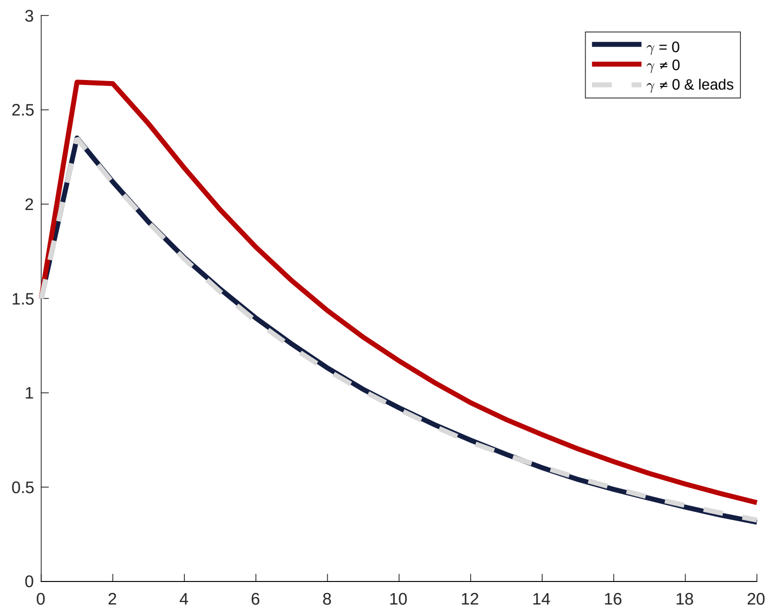
Note that equation (10) must include a lag of shock x_t to capture the effect of δ_1 in system (2). However, this does *not* control for the potential persistence of shock x_t, as will be apparent in the simulations.

Figure B1 shows the results of our simulations. In case (i) (dark-blue solid line), the response has a contemporaneous effect of $\hat{\delta}_{0,0} = 1.5$ and peaks at the following period due to the the fact that both ρ and δ_1 have positive values. Using the language from Section 3, the impulse response function estimated by LPs with no persistence is asymptotically equivalent to the one obtained directly from equation (10), that is, $\hat{\mathcal{R}}(h)^{LP} \to \mathcal{R}(h)^*$.

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 sets $\delta_{1+,h} = 0$, then the 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

[1] The choice of $\gamma = 0.2$ is based on an empirical application presented in Section 4. Of course, larger values of γ would yield higher biases due to the persistence of the process.

Figure B1: 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 (10) assuming $\gamma = 0$ in equation (2). 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 (10) and still assuming $\gamma = 0.2$.

between both responses (blue and red lines in Figure B1).

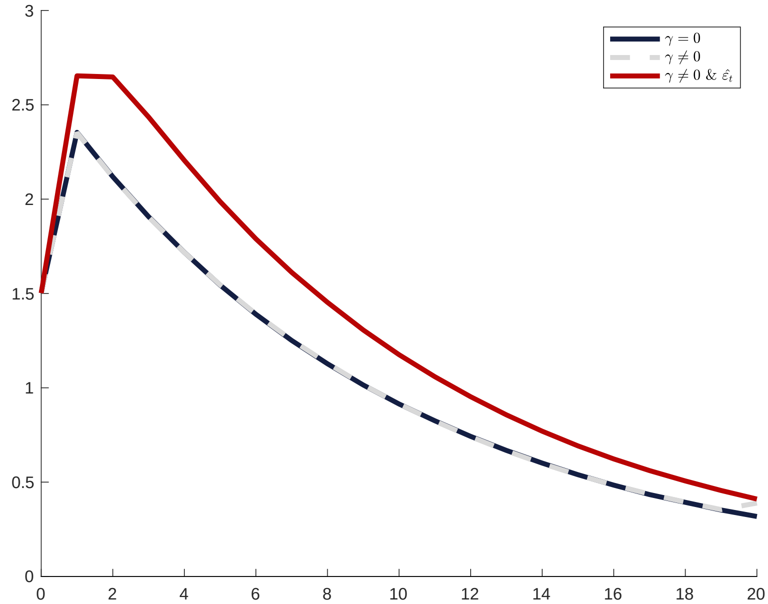
In case (iii) (dashed grey line in Figure B1), we see that the inclusion of leads of x_t renders the response of the outcome variable to a persistent shock identical to the one obtained when considering a shock without persistence, i.e., $\hat{\mathcal{R}}(h)^F \to \mathcal{R}(h)^*$. In Appendix B.2 we provide an alternative simulation where the shock x_t in (2) is not assumed to follow an AR(1) process but it is instead taken from actual data.

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

First, note that, since $\rho < 1$, system (2) can be inverted and re-written as:

$$y_t = (1 - \rho L)^{-1} (D_0 + D_1 L) x_t + (1 - \rho L)^{-1} u_t, \tag{B.1}$$

Figure B2: Simulated responses using DLMs



This figure shows the response of a simulated outcome variable to a shock with different degrees of persistence, using DLMs . The dark blue line shows the results of estimating equation (5) assuming $\gamma = 0$ in system (2). The dashed grey line shows the same estimation when $\gamma = 0.2$. The red line shows the response when substituting x_t in equation (5) by $\hat{\varepsilon}_t$, an OLS estimate of ε_t (see equation (9)), where serial correlation has been removed.

where L represents the lag operator.

Given the independence of u_t and x_t, the representation from equation (B.1) suggests that the dynamic responses of y_t from x_t can be obtained from the coefficients ϑ_h in equation (5).[2]

We estimate equation (5) fo three different cases: (i) assuming that $\gamma = 0$ in the DGP described in system (2), (ii) assuming that $\gamma = 0.2$ and (iii) replacing x_t with $\hat{\varepsilon}_t$ in equation (5) (i.e., following equation (12)).

The results are shown in Figure B2. Cases (i) and (ii) are displayed in blue and dashed grey lines, respectively. As argued earlier, since equation (5) 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 $\hat{\mathcal{R}}(h)^{DLM} \rightarrow \mathcal{R}(h)^*$ for any γ. Note that these impulse response functions are the same as

[2]Baek and Lee (2020) show that for autoregressive distributed lag models, setting the lag order to H is a necessary condition to achieve consistency.

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 B2. As argued in the previous subsection, when computing the impulse response with respect to ε_t, we are allowing the DLMs 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 DLM, we are not taking into account the persistence of x_t. In this case, the responses are equal to those obtained from LPs when $\gamma \neq 0$: $\hat{\mathcal{R}}(h)^{DLM-per} = \hat{\mathcal{R}}(h)^{LP} \rightarrow \mathcal{R}(h)$.

B.2 Alternative example: 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:

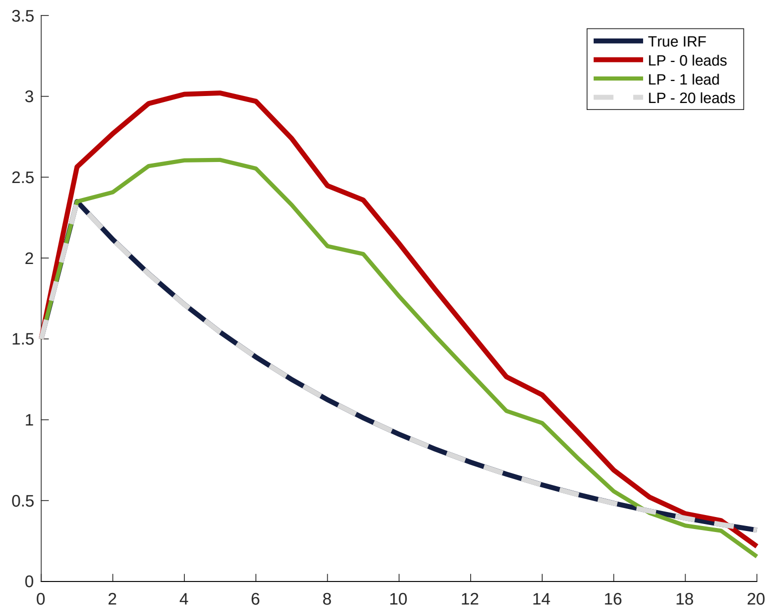
$$y_t = \rho y_{t-1} + B_0 x_t + B_1 x_{t-1} + u_t, \tag{B.2}$$

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.2) 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 B3.

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 red and dark blue lines in Figure B3). 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

Figure B3: 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.

(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).

B.3 Responses in local projections 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 (9). Then, ε_t can be used as substitute of the original shock x_t.

Assuming $\delta_1 = 0$ in system (2) (for simplicity) consider the following series of LPs:

$$y_{t+h} = \rho_h y_{t-1} + \lambda_h \varepsilon_t + \xi_{t+h}. \tag{B.3}$$

To obtain the dynamic responses of y_t to the shock ε_t (adjusted for persistence), we rewrite the first equation in system (2) as a function of ε_t and compute the relevant partial derivatives. For the cases of $h = 0$ and $h = 1$ these are:

$$\begin{aligned} \lambda_0 = \frac{\partial y_{t+1}}{\partial \varepsilon_t} &= \delta_0 \\ \lambda_1 = \frac{\partial y_{t+1}}{\partial \varepsilon_t} &= \rho \frac{\partial y_t}{\partial \varepsilon_t} + \delta_0 \frac{\partial x_{t+1}}{\partial \varepsilon_t} = \rho\delta_0 + \delta_0\gamma = \delta_0(\gamma + \rho). \end{aligned} \tag{B.4}$$

That is, even after correcting for the persistence in shock x_t, conventional LPs yield responses $\mathcal{R}(h)$, i.e., still containing the effect of persistence of the shock.

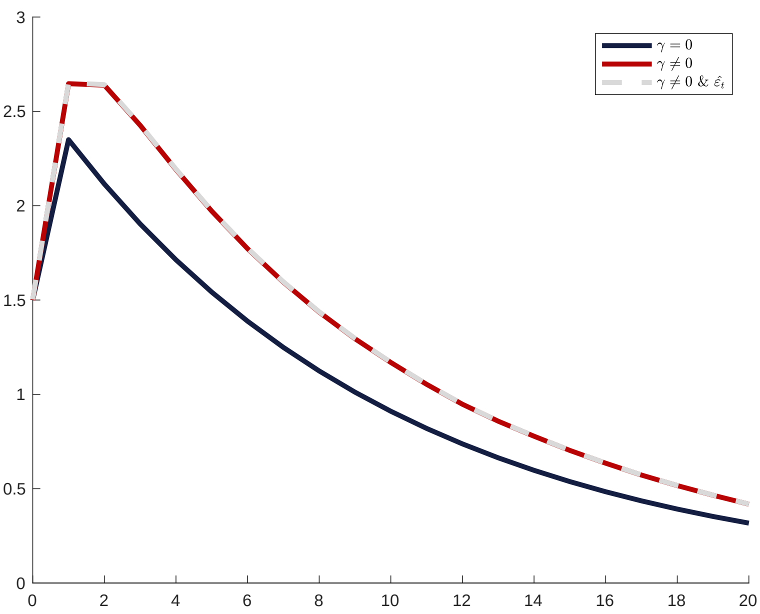
While this result may seem counter-intuitive, it arises from the fact that LPs do not have an explicit dynamic structure as a DLM. Hence, removing the persistence from x_t does not eliminate its effect on y_{t+1}, y_{t+2}, etc.

To empirically show this point, we simulate series of y_t and x_t following system (2) and the calibration used in Section B.1 (we now allow $\delta_1 \neq 0$). We then obtain the residuals $\hat{\varepsilon}_t$ as an estimate of ε_t described above and estimate the following equation:

$$y_{t+h} = \rho y_{t-1} + \lambda_{h,0}\hat{\varepsilon}_t + \lambda_{h,1}\hat{\varepsilon}_{t-1} + \xi_{t+h}. \tag{B.5}$$

Results are shown in Figure B4. 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.3) fails to retrieve an impulse response as the one obtained when $\gamma = 0$ in equation (2).

Figure B4: 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.5) assuming $\gamma = 0$ in equation (2). 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.3)).

C Additional empirical applications

C.1 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.[1] As mentioned in Table 1, this measure of fiscal changes exhibits some degree of persistence.[2]

To explore the effects of persistence in this context, we compute the responses estimating a series of LPs:[3]

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

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, and $\boldsymbol{X}_{it}$ 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 (C.1) by setting $\beta_{h,f} = 0 \ \forall h, f$. The results, shown in black solid lines in Figure C1 qualitatively replicate the benchmark results of Guajardo et al. (2014), with a fiscal consolidation shock significantly reducing output during the first 6 years.[4]

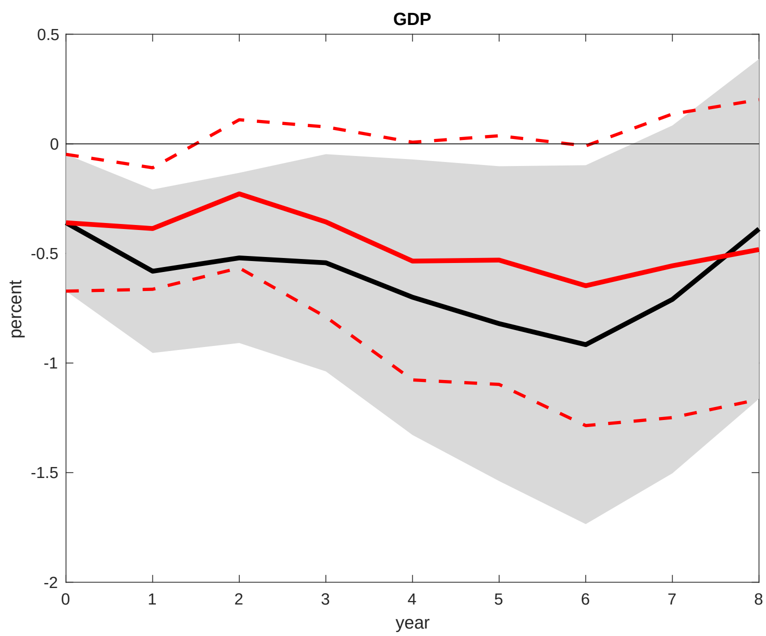
[1] A detailed description of these shocks can be found in Devries et al. (2011).

[2] 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).

[3] 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 et al. (2022) or Goujard (2017)).

[4] 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 D5 in Appendix D. As

Figure C1: Output response to a fiscal consolidation shock, with and without leads



Black lines show the results from equation (C.1) 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 (as 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.

Next, we estimate equation (C.1) but allow $\beta_{h,f} \neq 0$ (red lines in Figure C1). 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 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.

This exercise suggests that the policy implications from fiscal consolidations may be different when estimating $\mathcal{R}(h)$ *vs.* $\mathcal{R}(h)^*$.

C.2 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

in the original paper, we also find a significant reduction in this variable during the first 6 years after a consolidation shock.

on the estimation of dynamic responses. Beyond corroborating the previous statement, this subsection shows that the unnecessary inclusion of leads does not negatively affect inference in this application.

We estimate the response of output to exogenous tax changes following Romer and Romer (2010). We adapt the original estimation from the authors to the LPs setting:[5]

$$\frac{y_{t+h} - y_{t-1}}{y_{t-1}} = \beta_{h,0} shock_t + \sum_{f=1}^{h} \beta_{h,f} shock_{t+f} + \xi_{t+h}. \tag{C.2}$$

In our first exercise, we set $\beta_{h,f} = 0 \ \forall h, f$ in equation (C.2) to replicate the results from Romer and Romer (2010). The results are shown Figure C2 (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.[6]

Next, we allow for $\beta_{h,f} \neq 0$. The results, shown in Figure C2 (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.

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 C2 shows that confidence bands are not distinguishable between both specifications during the first seven quarters and differ only slightly afterwards.

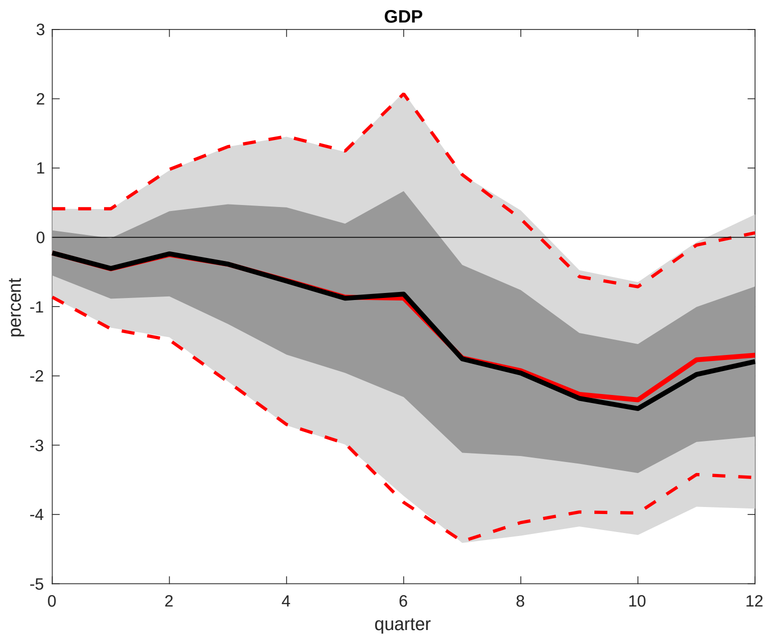
In sum, this placebo exercise is reassuring in that the inclusion of leads only matters when the explanatory variable displays some persistence. These results suggest that including leads in LPs is a conservative way to address the effects of persistence when there is a suspicion that the shock is persistent and the researcher wants to identify $\mathcal{R}(h)^*$.[7]

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

[6] 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.

[7] See Alloza and Sanz (2021) for another example that adds leads to LPs using a non-persistent shock. Similarly to the evidence provided in this section, they also show that adding leads does not affect inference.

Figure C2: 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(C.2) 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 (C.2) 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.

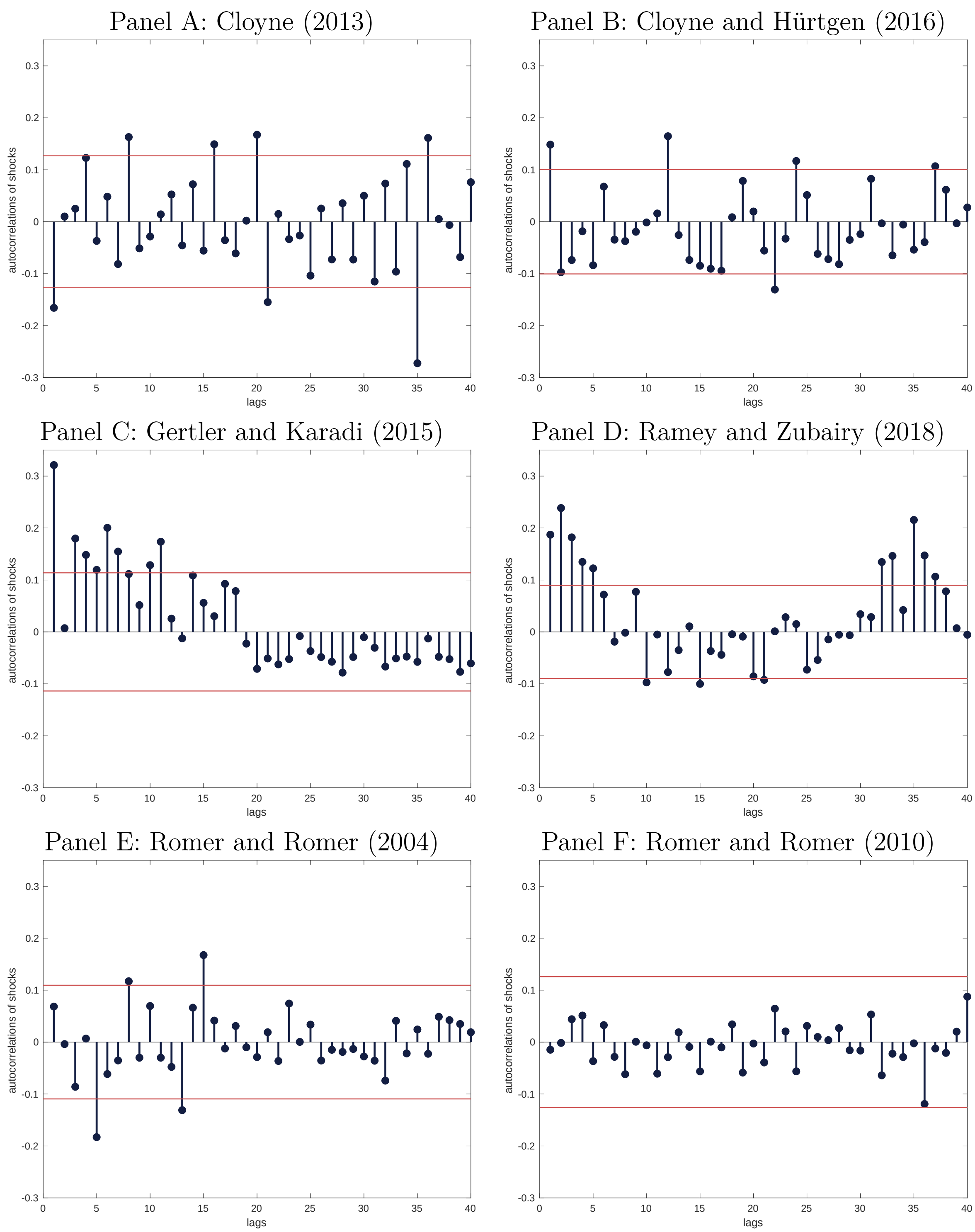
D Additional Tables and Figures

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



Red lines denote 95% confidence intervals.

Figure D2: Time series of shocks

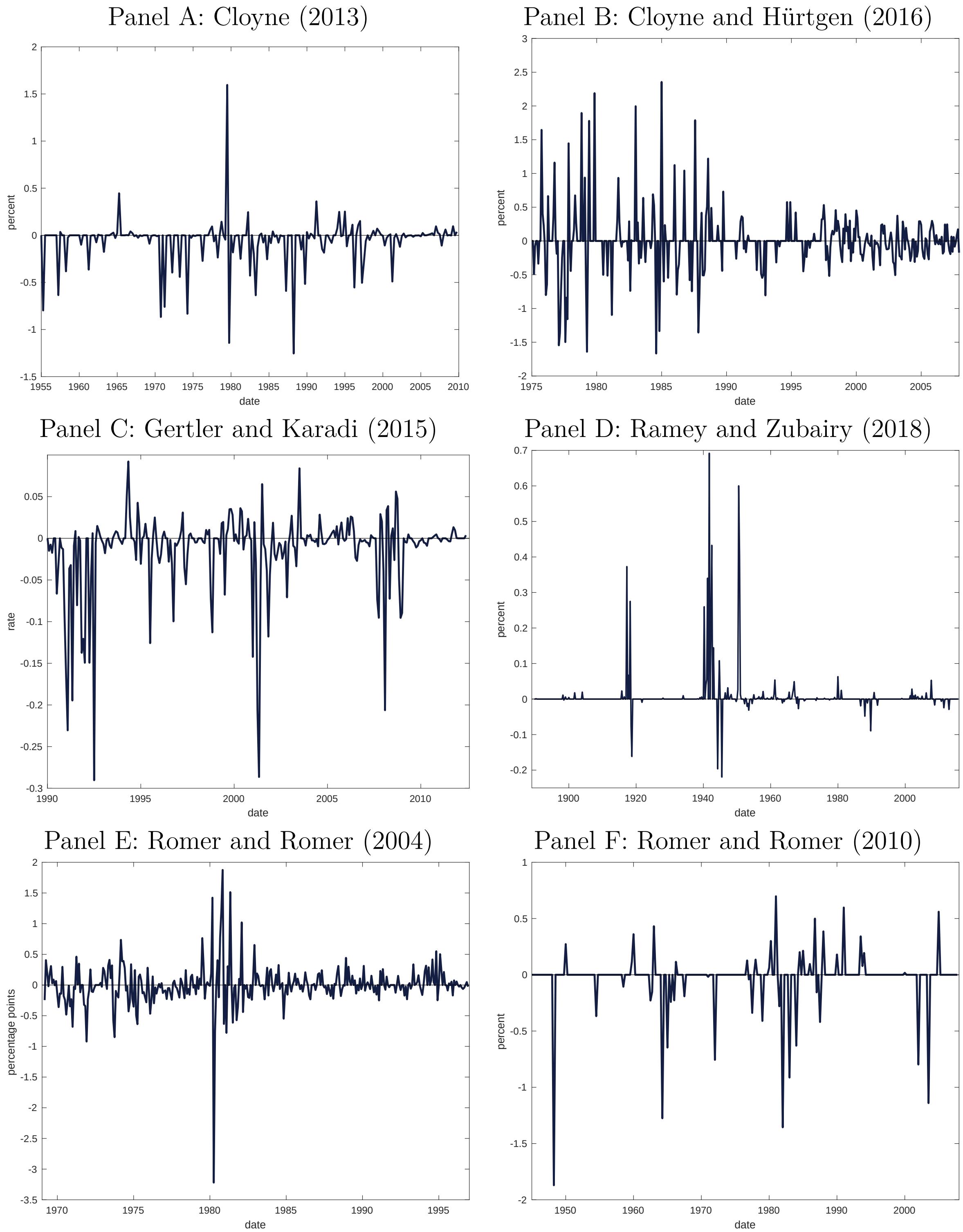
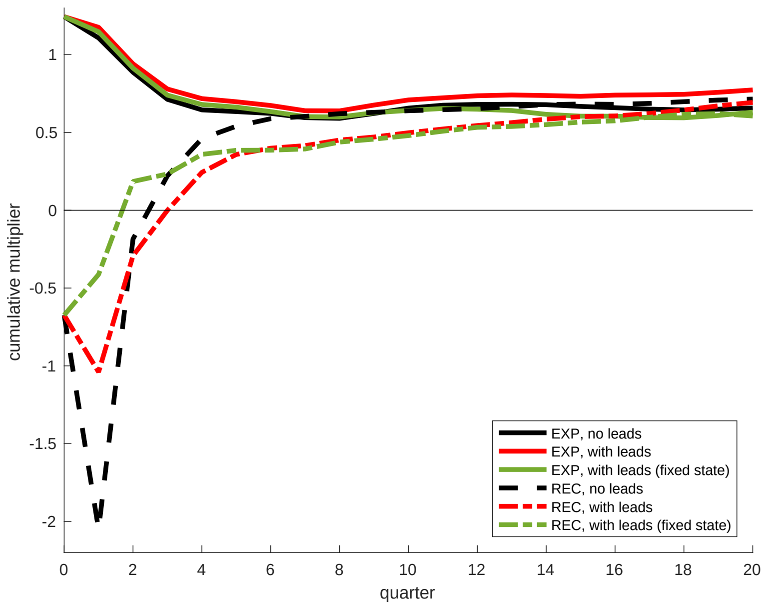
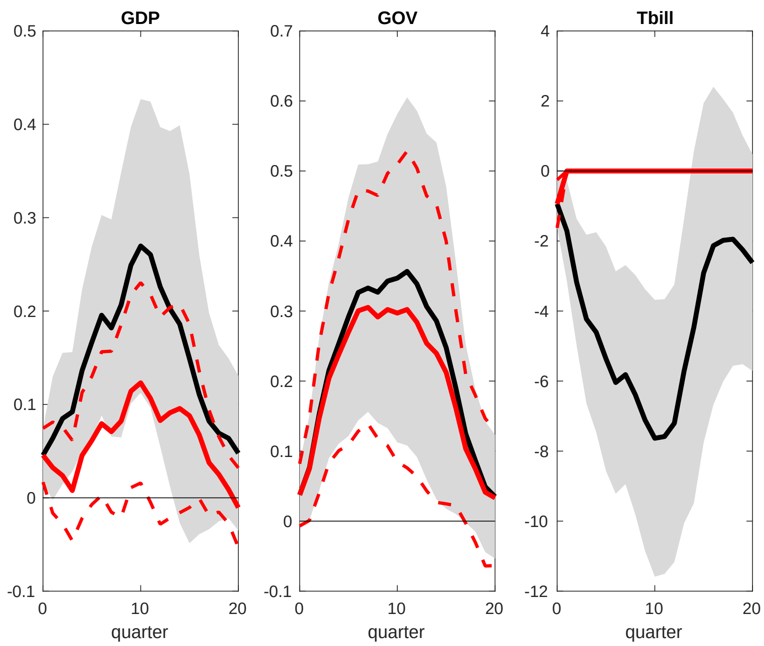


Figure D3: Government spending multiplier during expansions and recessions, with and without leads



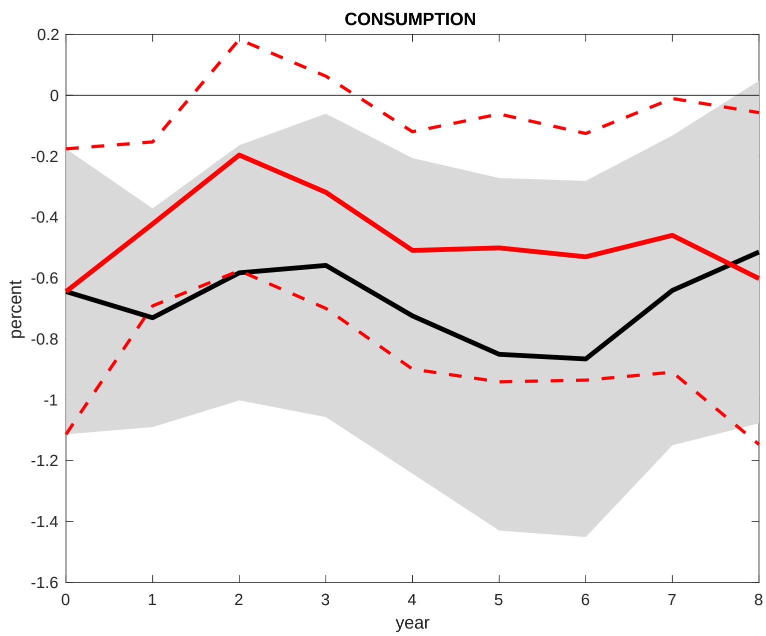
The black solid and dashed lines show the cumulative multiplier during periods of expansion and recession, respectively, without including any lead (as in Ramey and Zubairy (2018)). The red solid and dashed lines show the cumulative multiplier during periods of expansion and recession, respectively, 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 D4: Output and government spending responses, with and without leads of the unemployment rate



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

Figure D5: Private consumption response to a fiscal consolidation shock, with and without leads



Black lines show the results from equation (C.1) 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.