OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12646

Long-run Effects of Austerity: An Analysis of Size Dependence and Persistence in Fiscal Multipliers

GUILHERME KLEIN MARTINS

Department of Economics, University of Leeds, Leeds, UK (e-mail: g.kleinmartins@leeds.ac.uk; guilherme.klein.martins@gmail.com)

Abstract

This paper provides evidence that austerity shocks have long-run negative effects on GDP. Our baseline results show that contractionary fiscal shocks larger than 3% of GDP generate a negative effect of more than 5.5% on GDP even after 15 years. Evidence is also found linking austerity to smaller capital stock and total hours worked in the long-run. The results are robust to different fiscal shock datasets, the exclusion of particular shocks, and the use of cleaner controls. The paper also engages with the emerging discussion regarding fiscal multipliers heterogeneity, presenting evidence that the effects of exogenous fiscal measures are nonlinear on the shock size. The results also contribute to the broader discussion on the long-run effects of demand by suggesting that such shocks might permanently affect the economy.

> *(...) macroeconomics in this original sense has succeeded: Its central problem of depression prevention has been solved, for all practical purpose (...) the potential for welfare gains from better long-run, supply-side policies exceeds by far the potential from further improvements in short-run demand management.*
>
> Lucas Jr. (2003)

> *This post-crisis experience suggests that changes in aggregate demand may have an appreciable, persistent effect on aggregate supply – that is, on potential output.*
>
> Yellen (2016)

I. Introduction

In August 2022, Greece exited the European Union's 'enhanced surveillance', a framework established to ensure the policies implemented in the country from 2010 would not be reversed. These measures, aimed at decreasing public indebtedness, included large cuts

JEL Classification numbers: C54, E62, H5, H2.
I thank Daniele Girardi and Laura Carvalho for critical suggestions to this paper. I also thank Ethan Ilzetzki, Yuriy Gorodnichenko, Peter Skott, Arslan Razmi, Christian Rojas, and two anonymous referees for helpful comments on earlier versions. All remaining errors are mine.

to public spending, privatizations, and tax increases. After 12 years of its implementation, it is not clear how successful the strategy was. Greece's general government debt went from 130% of GDP in 2010 to 224% in 2021, while the average of OECD countries went from 70% to 94.7%. Greek real GDP *per capita* in 2021 is still 12.7% lower than in 2010, while the European Union (EU) expanded 12.1%.[1] The labor market was also impacted significantly: while the EU had an increase of 3.5% in its labour force, Greece had a reduction of 8.4%. Moreover, long-term unemployment[2] increased by more than 41% in the country between 2010 and 2021, while it fell by 7% in OECD.

Ideally, to evaluate the success of the austerity strategy, one would have to compare Greece's performance in the period to what would have happened if different policies had been implemented. Moreover, to take more general conclusions that can inform policy, it is also relevant to understand the timing of effects; that is, how much of the decrease in GDP in 2021 is related to the austerity implemented in 2017 and how much to the policies applied still in 2010, for instance. Such analysis of the long-run effects of austerity, however, is nonexistent in the literature, despite being central to the discussion that dominated economic and policy debates in recent decades. This paper seeks to fill this gap.

In different moments in the past 15 years, due to economic crises, such as the financial in 2007, the debt one in the Eurozone, and the Covid pandemic, or by broader theoretical reasons, such as the discussions of a 'secular stagnation' and a zero-lower bond for monetary policy, more aggressive fiscal policy has been brought to the fore. This movement has also been accompanied by a 'renaissance in fiscal research', as pointed out by Ramey (2019), which led to a significant improvement in our knowledge about the topic. The literature, however, focuses on (i) the short and medium-runs effects of (ii) fiscal shocks in general.

There might be different reasons for the shorter-run focus. Ramey (2019) points to methodological issues, arguing that the methods to estimate long-run effects would be different than those commonly employed in the fiscal literature. Another potential explanation is the theoretical understanding that demand shocks have only short-term effects, with supply-side factors determining the long-run. Both arguments, however, should not prevent an interest in estimating the long-run effects of these shocks. First, there are now methods widely used in the literature to estimate the effects of similar shocks over extended time horizons. Second, although the idea of neutrality of demand in the long-run is still important, there has been growing interest in recent years in the long-term effects of shocks, particularly negative ones related, for instance, to political, banking, or financial crises (e.g. Yellen, 2016 and Blanchard *et al.,* 2015). By estimating the long-run effects of fiscal shocks, one can also contribute to this emerging literature on the persistence of demand shocks.

Not least important is the fact that the literature tends to analyse the effects of fiscal shocks in general, and not of austerity policies. This is not only an important gap but, not rarely, a source of misunderstanding as the estimated effects of fiscal shocks in general are implied to hold for austerity measures in particular. Due to its deep implications

[1] Data from the World Bank. Calculated at 2015 constant US dollars.

[2] As a share of total unemployment. Long-term unemployment defined as unemployment by more than 1 year.

on social and political spheres, economists' use of the term should dialogue with other areas of knowledge and the broader public, for which austerity tends to mean 'enforced or extreme economy especially on a national scale', as defined by the Merriam-Webster dictionary. Less anecdotally and without resorting to other fields, this is also recognized by Alesina *et al.* (2019a): '[t]he term ''austerity'' indicates a policy of *sizeable reduction* of government deficits and stabilization of government debt achieved by means of spending cuts or tax increases, or both.' (p. 1, italics added). The literature, however, with very few exceptions,[3] ignores this definition in important ways; of particular interest for us in this paper, is the assumption that the effects are linear on the size of the shock.[4]

There are different reasons why the size of the shock might be relevant. The economy can have multiple equilibria,[5] and shifts between equilibria might depend on the size of the initial departure from the former steady state. Another channel could be through to money illusion, for instance, and the idea reinforced by behavioural research that the cognitive costs related to operating with nominal or real values are nonlinear (e.g. Fehr and Tyran, 2001). The effects might also be asymmetrical due to different reactions of the financial market: as in Greenwald *et al.* (1988), banks are resilient to relatively small negative shocks, but sufficiently large ones can lead to financial collapse. An additional channel might be related to factor hoarding: in face of a small demand shock, output might be adjusted via changes in capacity utilization and work intensity, while larger shocks tend to generate modifications in investment plans and labor demand, with larger impacts on aggregate demand. All these reasons might impact not only the proportional effects (or multiplier, in a more general usage of the term) of the shocks, but also their persistence over time. Therefore, taking into consideration the size of the fiscal shock is important not only as a matter of following the definition of austerity, but also because there are multiple theoretical reasons indicating that the effects might not be symmetrical and proportional.

There is a flourishment of research on asymmetrical and nonlinear effects of macroeconomic shocks. Regarding monetary policy shocks, some examples are Tenreyro and Thwaites (2016), Barnichon and Brownlees (2019), and Stenner (2022) that test for state, sign and size-dependency. Recent examples on the fiscal policy side are Barnichon *et al.* (2022), which use US data and conclude that, within a 5-year horizon, the multiplier of negative shocks is larger than 1, while the multiplier for expansionary shocks is always smaller than 1, and Ben Zeev *et al.* (2023) that, on the other hand, also using US data, finds no difference in the magnitude of multipliers of positive and negative shocks. These papers, however, do not look at the key aspect of asymmetry that we are considering here: the size of the shock. The discussion of how the empirical literature deals with these dimensions is resumed in section II.

[3]Recent papers by Ben Zeev *et al.* (2023) and Barnichon *et al.* (2022) explore the issue of symmetry in fiscal multipliers of contractions and expansions, for instance. Alesina and Ardagna (2010) is also an important exception, as the authors calculate separatedly the effects of expansions and contractions using the CAPB method.

[4]As will be resumed in section II, in some sense the size of the shock is relevant for an important strand of the literature, as in Alesina *et al.* (2019b), in which the size matters as the average elasticity is calculated; or in Alesina and Ardagna (2010), in which they declare a shock only changes in the adjusted primary balance larger than 1.5% – in this case, again, however, it is only the average effect that is calculated.

[5]Due to increasing returns to scale, or asymmetric information, for instance.

This paper aims to fill these important gaps in the literature by estimating the effects of austerity – understood as contractionary fiscal shocks of significant magnitude – over a time horizon of 15 years. Results indicate that sufficiently large shocks (more than 3% of GDP in the baseline case) generate a significant and persistent reduction in GDP even after 15 years; this result is robust to the use of alternative datasets, the exclusion of episodes, and the implementation of different estimation methods. We provide evidence that both the capital stock and labor input (proxied by total hours worked) are significant channels contributing to the long-term negative impact on GDP. The paper also presents evidence that fiscal multipliers are heterogeneous by shock size. While the negative shocks, in general, have an instantaneous multiplier of 0.07 and a long-run one of 0.51, fiscal shocks larger than 3% of GDP are associated with multipliers of 0.23 and 1.45, respectively.

Besides this introduction, the paper has three other parts. In section II, we present the current research on fiscal shocks to locate this paper in the broad literature and introduce, by comparisons, the methodology used in the empirical estimations. Section III explains the method and data in more detail and presents our baseline estimations. It is followed by section IV, in which a series of robustness checks and extensions are performed. Section V concludes the paper.

II. Fiscal Research

There are two main methods used to estimate exogenous policy changes at the country-level. A traditional method in the literature is the cyclically adjusted primary balance (CAPB) method (e.g. Blanchard and Perotti, 2002, Alesina and Ardagna, 2010). The idea is that, by calculating how much the components of the government budget change along the economic cycle, one can net this effect from actual government primary balance and thus check if the public sector is acting with a positive, negative, or neutral impulse in the economy. This method has received multiple criticisms. Romer and Romer (2010) point out that CAPB is affected by non-policy changes that might be correlated with other elements affecting output.[6] Another argument, which goes to the heart of the endogeneity concern, is that even if the CAPB method correctly indicates a discretionary policy change, its motivation might be related to cyclical fluctuations: governments might cut spending if inflation is increasing; social expenditure tends to increase in recessions, and so on (e.g. Devries *et al.,* 2011; Ball *et al.,* 2013).

An alternative to CAPB[7] that recently gained ground is the 'narrative approach'. This method tries to look directly at exogenous fiscal shocks, that is, changes in government expenditure or revenue that are not related to the business cycle. In the most recent and consolidated datasets, these shocks are identified by the analysis of official documents (congressional debates, speeches, budget documents, etc.) and consider as exogenous

[6]An example given by the authors (a similar argument is made by David and Leigh (2018)) is a stock market boom that raises cyclically adjusted revenues due to capital gains realizations but also correlates with other elements in the economy that will generate a future increase in output.

[7]There are other procedures that are similar in spirit to CAPB. Mountford and Uhlig, 2009, for instance, main identification strategy using VARs is imposing sign restrictions: for instance, the impulse response function of the government revenue (spending) will be positive for four quarters following a positive shock of the same variable and, even more important, that the shock is orthogonal to the business cycle and monetary policy.

the changes motivated by the goal of increasing long-run growth or reducing the budget deficit.[8]

This method is increasingly recognized as an important step in improving estimations based on panel data. However, it is also not exempt from criticism. Jordà and Taylor (2016) show that the time of fiscal shocks in the IMF fiscal narrative dataset (Devries *et al.* (2011)) can be predicted by some state variables – for instance, fiscal consolidations are more likely when public debt to GDP is high and when GDP growth is below potential. They propose using a propensity weighting strategy to further improve the identification of fiscal shocks. More details of the method will be presented in section III.

In terms of methods to get impulse response functions of the output after the fiscal shocks, there are two main alternatives in the literature. The one used in this paper is based on Local Projections (Jordà (2005)), which has the advantage of not requiring the assumption of any particular functional form.[9] An alternative econometric method that is also widely used is Vector Autoregressions (VARs); it requires, however, the assumption of a model and, although generating a smaller variance, it tends to produce a more biased estimation, increasingly so for long horizons (Li *et al.* (2024); Jordà *et al.* (2020)).[10]

We can return to the observation by Ramey (2019), mentioned in section I, that the long-run effects of fiscal shocks are not estimated due to methodological limitations. Semi-parametric methods have been used in estimations with similar setups over long time horizons. Jordà *et al.* (2020) use local projections with instrumental variables to calculate the effects of monetary shocks over 12 years, and Acemoglu *et al.* (2019) implement local projections with different propensity weighting methods to estimate the effects of democracy on a 30-year horizon, to name a couple. Therefore, it is not unusual in recent research to use the methods implemented here to calculate long-run effects of similar shocks.

As also indicated in section I, another potential explanation for the lack of research on the long-run effects of fiscal shocks, however, is the theoretical understanding that demand shocks only have short-run effects, with supply determining the long-run. This view has prevailed in economic theory (Yellen, 2016), from 'standard' growth models, such as Solow (1956), to both new classical (and real business cycle) and most of the new Keynesian models, and has largely informed macroeconomic empirical research.[11] In recent years a number of papers resumed the discussion about the long-term effects of

[8]Another implementation adopted by the literature with this method is to look at military spending related to foreign conflicts (e.g., Ramey and Shapiro (1998)).

[9]Jordà and Taylor (2016) argue that the method also provides better control for observable variables and is more reliable when the instrumental variables (for the fiscal shocks) themselves might be endogenous.

[10]Ramey and Zubairy (2018) use the paper of Auerbach and Gorodnichenko (2012) to exemplify other differences between using local projections (LP) and VARs in those estimations, particularly in the context of estimating the effects of fiscal changes based on different states of the economy. With the Jordà method (Jordà, 2005), the transition between states (booms and recessions, for instance) appears directly if it is caused by the (average) shock or is captured by the other control variables. With regime-switching VAR models, as in Auerbach and Gorodnichenko (2012), one has to make assumptions; in this case, about when the parameters should switch between states (they assume that economic states last for at least 20 quarters). In their subsequent work, Auerbach and Gorodnichenko (2013) perform a very similar exercise, but using local projections instead of structural vector autoregression due to the advantage mentioned above, but also because local projections tend to facilitate the correction of errors correlation within countries and it does no constrain the shape of the IRF.

[11]A classical example is Blanchard and Quah (1989).

negative shocks, but most focus on the effects of political, banking, or financial crises, while others look at GDP and estimations of its trend to identify recessions and analyze their effects over time (Haltmaier, 2013 over a 4-year horizon; Cerra and Saxena, 2008, Martin *et al.,* 2015 and Blanchard *et al.,* 2015 over a maximum horizon of 10 years are some examples).

However, there are very few estimations for fiscal shocks. The main exception[12] is the emergence of a literature in recent years analysing the long-run effects of public R&D expenditure. Li and Koustas (2019) use a 'quasi-experimental' framework, and estimate the long-run effect of US Second World War defense spending in local economies. Gross and Sampat (2023) also explore public investment in the US during World War II to analyse effects at the regional level. Kantor and Whalley (2023) estimate the effects on counties that received more public R&D conducted by NASA contractors during the Cold War era Space Race. The work of Antolin-Diaz and Surico (2024) is the closest paper to ours in terms of the goal of estimating the long-run effects of government spending. The authors use the dataset of military spending news constructed by Ramey and Zubairy (2018) as the government spending shock to test the effects on different macro variables on a fifteen-year horizon.

These papers advance the literature greatly but differ from ours in important dimensions. First, they focus on specific government expenditure types rather than general fiscal policy. Second, they estimate the effects in a single country. Third, the ones focused on R&D rely on a 'quasi-experimental' framework and analyse the effect of a policy in one specific time frame; in the case of Antolin-Diaz and Surico (2024), concerns are if the results can be extended to any other country, given the particularity of the military sector in the US economy. Thus, there are important concerns about external validity issues in both cases. Lastly, in the case of Antolin-Diaz and Surico (2024), the exogeneity assumption implied in their baseline Cholesky decomposition is based on a weak instrument, at least after 1954 (Ramey, 2011).

Table 1 lists some of the most influential papers in the fiscal research literature. The literature is vast, and it is difficult to do justice to all the important contributions; the list is produced to include more recent papers closer to ours in estimating shocks using country-level data and to illustrate the diversity of empirical methods used. In terms of the results, the literature is also heterogeneous, although there has been a convergence in recent years towards the direction of the short-run effects on GDP of fiscal consolidations to be negative (Ramey, 2019). There are important exceptions, however, such as Giavazzi and Pagano, 1990 and more recently Alesina and Ardagna, 2010, which argue that permanent contractionary fiscal shocks can have a positive effect on output, sparking an intense discussion around the 'expansionary austerity' hypothesis.[13]

[12]There is also a literature that analyses errors in GDP forecast to estimate the effect of fiscal shocks. Fatás and Summers (2018), for instance, focus on consolidations that took place in 2010–11 and whose estimations are completely based on forecasts, both for GDP (up to 2021) and for the structural balance. A similar example is Gechert *et al.* (2019), which uses narrative identified fiscal shocks.

[13]The authors propose a few channels through which the effects could take place. On the demand side, if agents believe that the shock prevents a much more disruptive adjustment in the future, it would generate a positive wealth effect, which might increase demand. Also, if agents believe the adjustment is credible and avoids default, they would ask for lower premiums on government bonds, reducing interest rates. On the supply side, the main channel would be via the labour market. Expenditure cuts (in government jobs and wages, for instance) would

TABLE 1

Selected studies of the effect of fiscal shocks on GDP

Authors	*Data*	*Identification*	*Method*	*Max. horizon*
Alesina *et al.* (2019b)	16 OECD countries 1978–2014	Narrative	VAR	5 years
Jordà and Taylor (2016)	17 OECD countries 1978–2009	Narrative	LP (AIPW)	5 years
Riera-Crichton *et al.* (2016)	15 OECD countries 1980-2009	Narrative (VAT changes)	LP	1 year
Guajardo *et al.* (2014)	17 OECD countries 1978–2009	CAPB inst. by narrative	2SLS and VAR	5 years
Ilzetzki *et al.* (2013)	44 countries 1960-2007	CAPB (Expenditure)	VAR	5 years
Baum *et al.* (2012)	6 OECD countries 1965–2011	CAPB	TVAR	3 years
Auerbach and Gorodnichenko (2012)	USA 1966-2009	CAPB inst. by forecast	STVAR	5 years
Romer and Romer (2010)	USA 1947–2007	Narrative (Tax)	OLS and VAR	5 years
Alesina and Ardagna (2010)	21 OECD Countries 1970–2007	CAPB	OLS	3 years
Mountford and Uhlig (2009)	USA 1955–2000	Sign restriction	VAR	6 years

Abbreviations: AIPW, augmented inverse propensity weighted estimator; STVAR, extension of smooth transition autoregressive (STAR) models; TVAR, threshold vector autoregression.

The most obvious difference between our estimation and the literature indicated in Table 1, as addressed at length, is the maximum time horizon.[14] However, there is another important element that is common in these works and, as mentioned in section I, is explored in this paper: the assumption of linearity of the effect of fiscal change, particularly that shocks of different sizes have the same proportional effects.

There are papers that deal with nonlinearities in the context of government spending. Besides the recent papers by Ben Zeev *et al.* (2023) and Barnichon *et al.* (2022), already mentioned, other important works that explore nonlinearities are, for instance, Auerbach and Gorodnichenko (2012), Ramey and Zubairy (2018), and Ghassibe and Zanetti (2022). However, they focus on state-dependency (differences in multipliers when the economy is in a recession or expansion)[15] or sign-dependency (differences in multipliers for positive and negative shocks), and not on size-dependency, as this paper does.

This linearity assumption tends to be more explicit in papers that use narrative fiscal shocks as the 'treatment' variable, given that not rarely the independent variable is binary. However, even in estimations with a 'continuous' treatment, that is, the size of the shock as

worsen workers' fallback position, decreasing wages in the private sector, allegedly increasing profits, investment and competitiveness.

[14]In some of the papers, such as in Ilzetzki *et al.* (2013), a 'long-run' effect is also calculated by assuming time goes to infinite; in practice, this is equivalent to the effect achieved with the convergence in the maximum horizon.

[15]Other studies also explore other variables linked to the state of the economy, such as debt levels, as in Iwata, and IIboshi, H. (2023).

TABLE 2
Description of narrative fiscal shocks in Alesina et al. *(2018)*

	Any size	*>1% GDP*	*>2% GDP*	*>3% GDP*
Total	232	128	69	33
Range of shocks	10.0%	8.7%	7.7%	6.7%
Avg. size of contraction	1.6%	2.5%	3.4%	4.4%

Notes: Range of shocks is the difference between the largest and smallest shock.

the independent variable,[16] for instance, a limitation persists, given that these estimations would still capture the *average* size of the effect, and, *a priori*, it is possible that shocks of different sizes have different proportional effects (or multipliers, in a more general use of the term). The limitation of taking into account only the average effects is highlighted in a sample with a large number of small shocks, which is the case for the narrative fiscal shocks datasets (see Table 2).

Summing up: our paper aims to address some gaps in the large and emerging fiscal research literature. The main one is to examine the long-run effects of austerity shocks, resorting to modifications in the estimation method to account for particularities of the time horizon and the fact that multiple shocks occur in the horizon of interest. Second, this paper does not assume that the fiscal multiplier is the same regardless of the shock size, which is particularly relevant not only to the conceptual discussion about austerity, but also to the emerging research on heterogenous fiscal multipliers.

III. Estimations

Baseline

As previously mentioned, despite its weakness, the narrative approach to identify fiscal shocks has been recognized in the literature as the best option to deal with endogeneity. In this paper, we use the dataset of narrative fiscal shocks compiled by Alesina *et al.* (2018), which is, to the best of our knowledge, the largest available, covering 16 OECD countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Portugal, Spain, Sweden, the UK, and the USA) from 1978 to 2014. The shocks identify permanent fiscal contractions that are exogenous to the economic cycle. Some basic descriptive statistics of the sample are displayed in Table 2; almost half of the contractionary shocks are smaller than 1% of GDP and only around 15% is larger than 3% of GDP.

For the reasons described in previous sections, we estimate the effects using a semi-parametric method. More specifically, we will use an extension of the Augmented Inverse Propensity Weighted Estimator (AIPW). According to Lunceford and Davidian (2004) and Jordà and Taylor (2016), the AIPW is the estimator with the smallest asymptotic variance within the class of the double-robust estimators – that is, those for which it is sufficient that either the conditional mean model ('outcome model') or the propensity score model ('treatment model') to be correctly specified for the estimator to be consistent.

[16]Which is the norm in estimations using VARs, but can also be applied with other methods, such as the Local Projections, as in Alesina *et al.* (2019b) and Riera-Crichton *et al.* (2016).

As indicated in section II, the 'treatment model' is used to calculate the probability of each unit (country-year) having an austerity shock. The variables used in the probit regression are:[17] country dummies, debt (% GDP), GDP gap,[18] real GDP growth (current and one lag), a dummy for an episode of fiscal consolidation in the previous year, long-term and short-term interest rates, current account (% GDP), change in the investment to GDP ratio, real private loan growth, and CPI inflation rate. Except for the data on the current account, which we extract from the OECD, and the one for real private loan growth, obtained with the Bank for International Settlements (BIS), the source for the other variables is the data employed by Alesina *et al.* (2018).[19] We use these probabilities of treatment in a propensity weighting strategy to further improve the identification of fiscal shocks: a higher weight is given to countries that, although having a higher probability of having a shock, do not have one.

As shown in the Appendix (sections A and B),[20] this procedure eliminates any differences between treatment and control groups with respect to the discussed variables and reinforces the exogeneity of the fiscal shocks. Figure 1 illustrates the effect of reweighting the sample. The fiscal shock occurrence distributions, initially significantly different for the control and treatment groups, become much more similar after the sample reweighting.

After the 'preliminary' stage of reweighting the sample, we can proceed to the 'outcome model', in which a regular difference-in-differences regression is performed with controls for conditional mean within a Local Projections set-up. We control for a cyclical component of GDP, country-fixed effects, and two lags of change in GDP.

More specifically, the estimator can be written as:

$$\begin{aligned}\hat{\Lambda}^{h}_{\text{AIPW}} &= \frac{1}{n}\sum_{t}\Bigg\{\left[\frac{D_t(y_{t+h}-y_t)}{\hat{p}_t} - \frac{(1-D_t)(y_{t+h}-y_t)}{(1-\hat{p}_t)}\right]. \\ &\quad - \frac{(D_t-\hat{p}_t)}{\hat{p}_t(1-\hat{p}_t)}\left[(1-\hat{p}_t)m^h_1(X_t,\hat{\theta}^h_1)+\hat{p}_t m^h_0(X_t,\hat{\theta}^h_0)\right]\Bigg\}.\end{aligned} \tag{1}$$

For which: y_{t+h} is the variable of interest at time $t+h$, D_t is the fiscal policy variable, $\hat{p}_t$ is the policy propensity score at time t given the relevant set of covariates contained at X_t, and m^h_j is a generic specification of the conditional mean of y_{t+h-y_t} in the subpopulation j (that is, with or without a shock). Finally, $\hat{\theta}^h_j = (\alpha^h_j \beta^h_j)'$, with α^h_j indicated what would

[17]As mentioned, this follows the procedure adopted by Jordà and Taylor (2016).

[18]The baseline estimations uses the widely used HP filter. Results are very similar if the GDP gap is measured by the Hamilton filter (Hamilton, 2018).

[19]It can be found here: https://www.igier.unibocconi.it/fiscalplans. The GDP data is in volume at market prices. For some data points, we had to make some minor adjustments. For four data points of indebtedness, we perform linear interpolation (Belgium 1989, Denmark 1997, Sweden 2003, Finland 1980). Moreover, for Germany and Ireland before 1990, we use the change in the correspondent variables of short and long-term interest rates in Jordà and Taylor (2016) to extrapolate these variables; the same procedure was implemented for CPI inflation in England before 1988 and for short-term interest rate from Sweden before 1982.

[20]We perform three exercises: (i) one shows that after reweighting, the control variables are not capable of predicting the treatment variable, (ii) a second makes use of the Durbin-Wu-Hausman test, (iii) and a third uses as the treatment variable only the part of the fiscal shock that is orthogonal to the control variables.

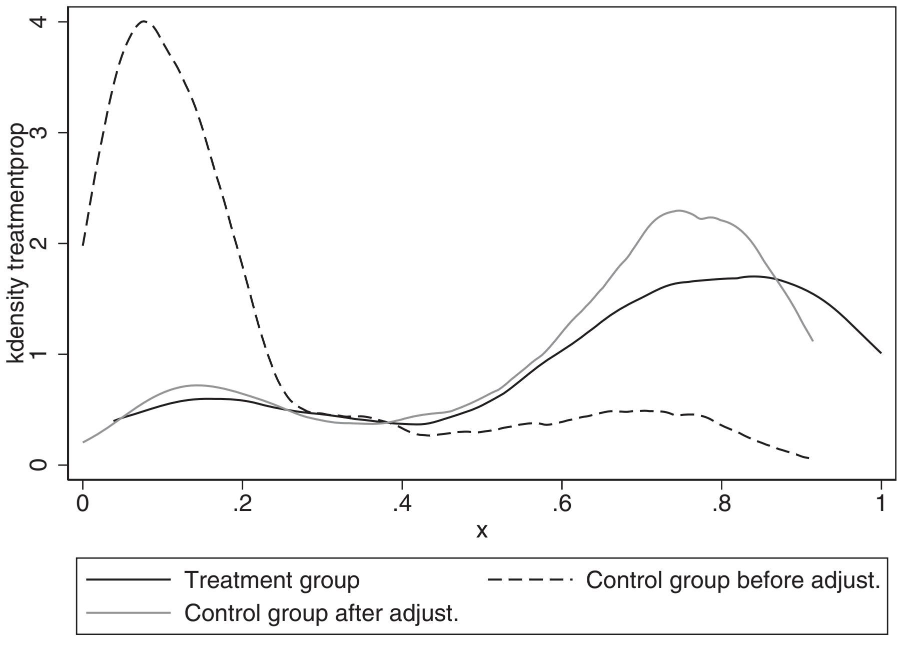

Figure 1. Fiscal shocks distribution probabilities.

be the size of $(y_{t+h} - y_t)$ for group j in the absence of treatment and β_j^h the estimator of the covariates over $(y_{t+h} - y_t)$.[21]

An important adjustment to this method is required. The main problem to be addressed here is that in settings in which the 'treatment' (austerity shocks) can occur multiple times, it is possible that, when interested in the effect of treatment at time t on $(y_{t+h} - y_t)$, another treatment takes place between time t and time h. In those cases, the effect of D_{t+j} for $j < h$ is absorbed by the fixed effects coefficients of the regression, biasing the estimation of the treatment itself. This problem increases with the forecasted horizon; thus, it is an important problem for long-run estimations such as the ones performed in this paper. The solution, proposed by Teulings and Zubanov (2014) and followed in this paper, is to include future fiscal shocks occurring up to time h in the future $(\sum_{j=0}^{h-1} \Lambda^h D_{t+h-j})$ as controls.

Figure 2 presents the main results of our estimations, namely the effects on GDP of contractionary fiscal shocks of different sizes. We are defining 'austerity' as negative shocks larger than 3% of GDP. We test other minimum thresholds in the robustness section (section IV).

As can be seen, when all contractionary shocks are considered, a negative effect on GDP is present in most years, but it is statistically significant only in the fourth and fifth years after the shock. The results are different for larger shocks: austerity episodes produce strong negative effects on GDP that are statistically significant for every year;

[21] In our baseline regressions, we will follow the assumption made in most macro estimations using VARs and which is also performed by Jordà and Taylor (2016) (table 8) that $\theta_0^h = \theta_1^h$.

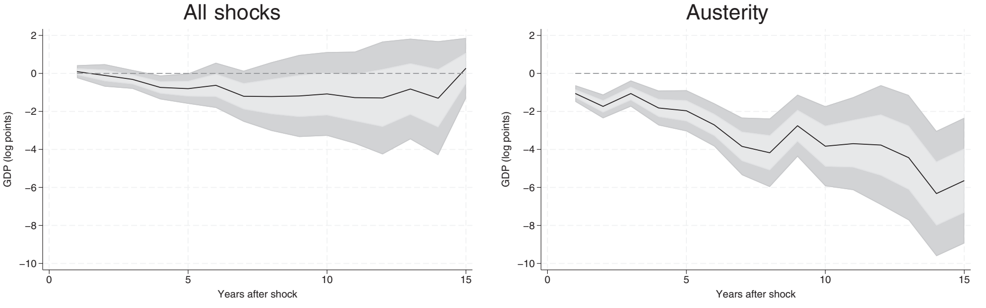

Figure 2. Effect of Austerity – By size of the shock
Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by two SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by one standard deviation (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP.

those shocks are associated with a reduction in GDP of 5.6% after 15 years.[22] In other words, our estimations suggest that relatively large contractionary fiscal shocks generate significant long-run negative effects on GDP.

Extended dataset and different GDP measure

The discussion regarding austerity regained centrality after the great financial crisis of 2007 and its repercussions in the European debt crises some years later. Given that our series goes up to 2014, an important limitation of the estimations is the exclusion of the long-run effects of this recent wave of austerity. A simple solution would be to extend the data on GDP; in our baseline specification, however, there is an additional problem: we are controlling for shocks occurring between t and $t+h$. Therefore, we would also need to extend the fiscal shock data.

Given the non-existence of a longer narrative dataset, we perform an intermediate solution: given the mentioned limitations of the CAPB method, we extend the series of shocks to be used only as controls with a measure of fiscal shock based on the cyclically adjusted primary balance calculated by the IMF, while we keep using the same narrative shocks as treatments. Following the usual procedure in the literature, we look at the annual change in the CAPB and assume that a shock occurs when the CAPB increases by at least 1.5% as a percentage of GDP. Finally, to generate a series for GDP up to 2019 – and to take into account that there might have been revisions in the growth rates since the data was compiled by Alesina *et al.* (2019a) – we use data from OECD on the growth rate of GDP (in volume).

As can be seen in Figure 3, qualitative results persist: for a sufficiently large austerity shock, there are statistically significant long-run effects on GDP.

[22]The control group is compounded by units (country-year) that do not go through an austerity shock. In Appendix C, we estimate using those units that do not have any fiscal constraints as controls. The results are very similar.

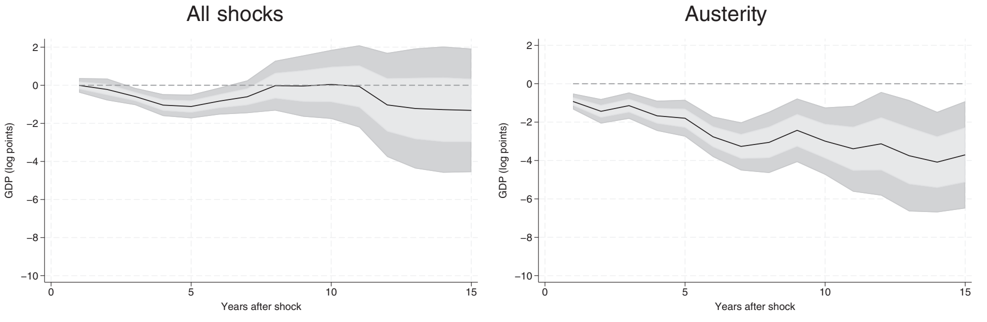

Figure 3. Effect of austerity – by size of the shock – extension
Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by two standard deviations (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP.

Fiscal multipliers and continuous treatment

We can now analyse the question of the proportional effects of different-size shocks. That is, if a shock 1% larger (as a % of GDP) has a different effect considering all the shocks and only those larger than 3%, for instance. This estimate gives us something similar to a fiscal multiplier.

To test this, we resort to an adaptation of our baseline method. First, in our 'treatment model', we reweight the sample the same way did before, using a binary treatment variable. In our 'outcome' model, however, we use a continuous treatment, that is, the size of the shock.[23] This is performed within each treatment band of interest of our baseline estimation: all contractionary shocks, and austerity ones (those larger than 3% of GDP).

Table 3 presents the results for the instantaneous and long-run 'multipliers'. The long-run coefficients indicate, for example, that a shock of 2.5% of GDP will, on average, reduce GDP by around 1.3% after 15 years, while a shock of 3.5% of GDP tends to produce a reduction of about 5.1%. The result indicates that the multipliers for sufficiently large shocks are significantly different than the one when considering all fiscal contractions: using a chi-square test, we can reject the hypothesis that they are statistically equal with a 5% significance level.

This result reinforces the idea that the size of the shock matters, not only due to persistence issues, as indicated in our baseline estimations, but also for potential non-linear proportional effects on the economy.

IV. Extensions and robustness

Alternative thresholds

To test how sensitive the results are to the austerity threshold used, we test the effect on GDP after 15 years of shocks considering four other thresholds. Figure IV indicates

[23] For the treatment itself and for its leads.

TABLE 3
'Multipliers' – by size shock

	All shocks	*Austerity*
Instantaneous (after 1 year)		
Multiplier	−0.07	−0.23
P-value ≥ 0% GDP	—	0.02
Long-run (after 15 years)		
Multiplier	−0.51	−1.45
P-value ≥ 0% GDP	—	0.03

Note: A Chi-square test is used to test the null hypothesis that the multiplier is equal to the one when all contractionary shocks are considered.

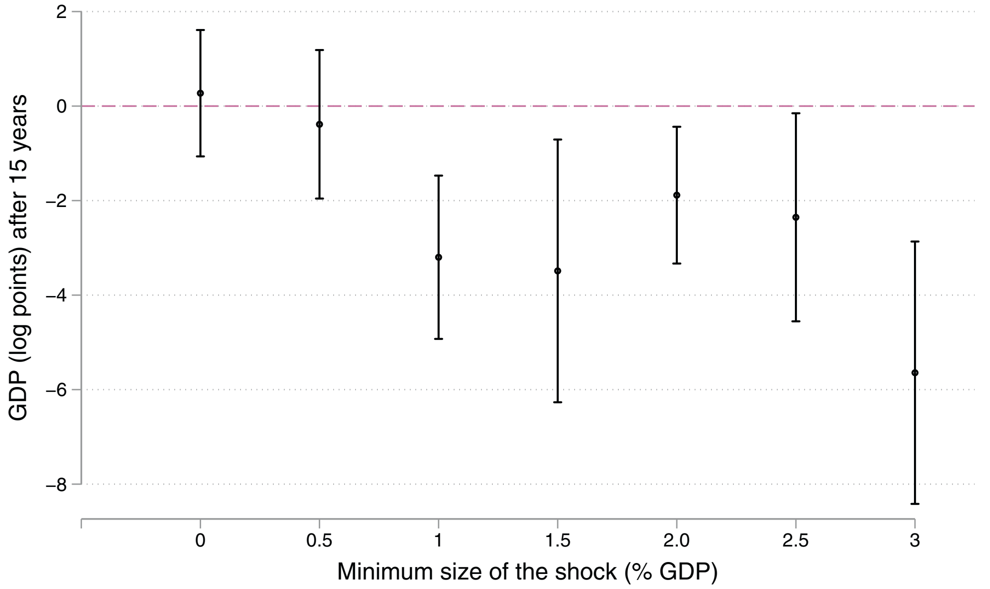

Figure 4. Alternative thresholds. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

that shocks larger than 1%, 1.5%, 2%, and 2.5% of GDP – besides the baseline cases of 3% – have long-run effects on GDP.

Alternative dataset

Another important dataset of narrative fiscal shocks is the one from Devries *et al.* (2011), which covers 17 OECD countries from 1978 to 2007. This dataset has several differences with respect to the one elaborated by Alesina *et al.* (2019a): the most explicit ones are the reduced number of years and the inclusion of the Netherlands. However, the changes are deeper, with frequent significant discrepancies in the size and timing of the shocks. Thus, checking if the effects of this alternative sample of shocks align with our baseline results can serve as an important robustness check.

The narrative shocks in the Devries *et al.* (2011) dataset tend to be smaller than the ones in Alesina *et al.* (2019a): the average shock size in the former is 1% of GDP, and the

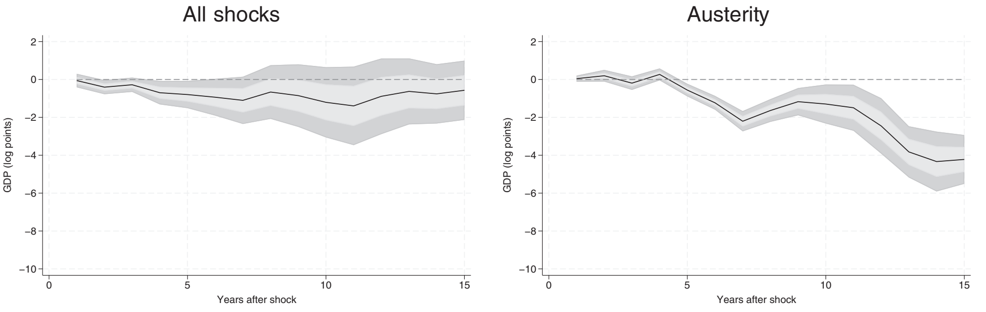

Figure 5. Effect by shock size – alternative dataset
Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by 2 SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP.

largest is 4.5%, while in the Alesina *et al.* (2019a) dataset, the average shock is of 1.6% of GDP, with the largest being of 9.7%. Therefore, there are not enough observations to perform the estimations for shocks larger than 3% of GDP, and thus, we define austerity shocks as those larger than 2.5% of GDP.

Figure 5 presents our results, which are very similar to the ones from Jordà and Taylor (2016), which use Devries *et al.* (2011) dataset, for short-run periods and considering all negative fiscal shocks, but for horizons longer than those estimated by the authors, the results are statistically insignificant. However, once again, when the shock size is taken into account, the results indicate something different. For austerity shocks, the negative effect on GDP is statistically significant from the fifth year onwards.

In Figure 6, we apply the same reasoning used in section IV and test different thresholds for the austerity definition. Again, we get a qualitatively similar result indicating that our findings regarding the long-run effects of austerity shocks are robust to different thresholds for the minimum size of the shocks.

Excluding episodes and countries

As indicated in Table 2, there is a wide spectrum of shock sizes, this being one of the key venues of exploration in our paper. However, given that we are placing only a lower limit to the shocks, particularly large austerity measures may be driving our results. To test the robustness of our results to this possibility, we re-run the baseline estimation for austerity shocks excluding one episode at a time and check if the effects on GDP after 10–15 years hold. Figure 7 shows that the results are robust to the exclusion of any particular shocks.

Another robustness exercise is to exclude entire countries from the sample. One reason for this exercise is the exclusion of a larger group of observations at each time (compared with the exclusion of particular shocks). Another is that it is possible that for some countries the shocks have a larger degree of endogeneity: for instance, contrary to Devries *et al.* (2011), Alesina *et al.* (2019a) exclude the Netherlands from their sample given that the fiscal rule of the country leads to a particularly large correlation between fiscal adjustments and past output growth. As can be seen in Figure 8, the result that there are long-run effects of austerity shocks (contrary to the case with all fiscal shocks) holds

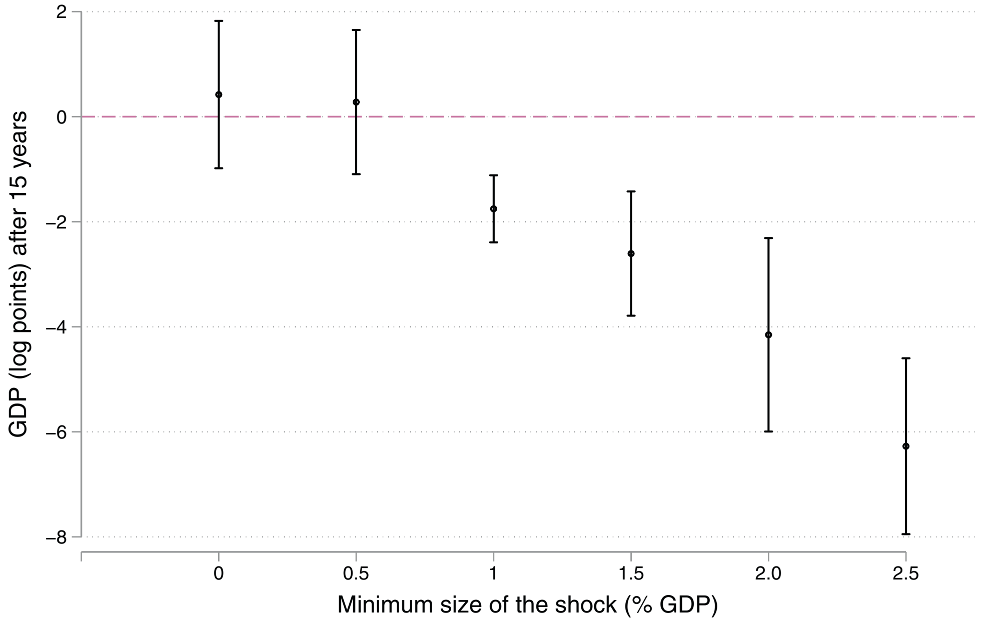

Figure 6. Effect by shock size – alternative dataset and thresholds. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

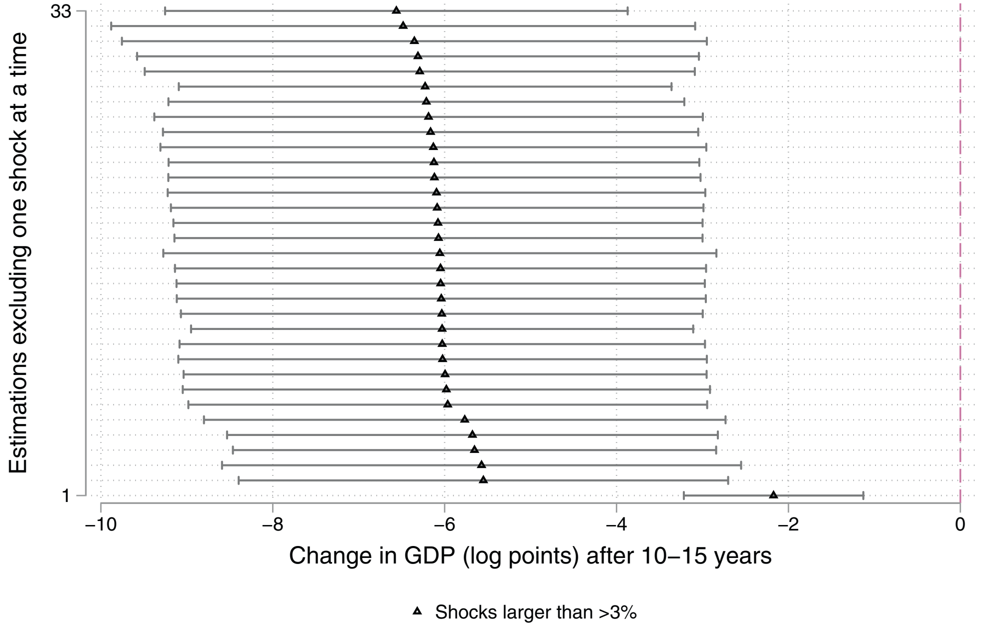

Figure 7. Robustness check – excluding shocks (larger than 3% of GDP). [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

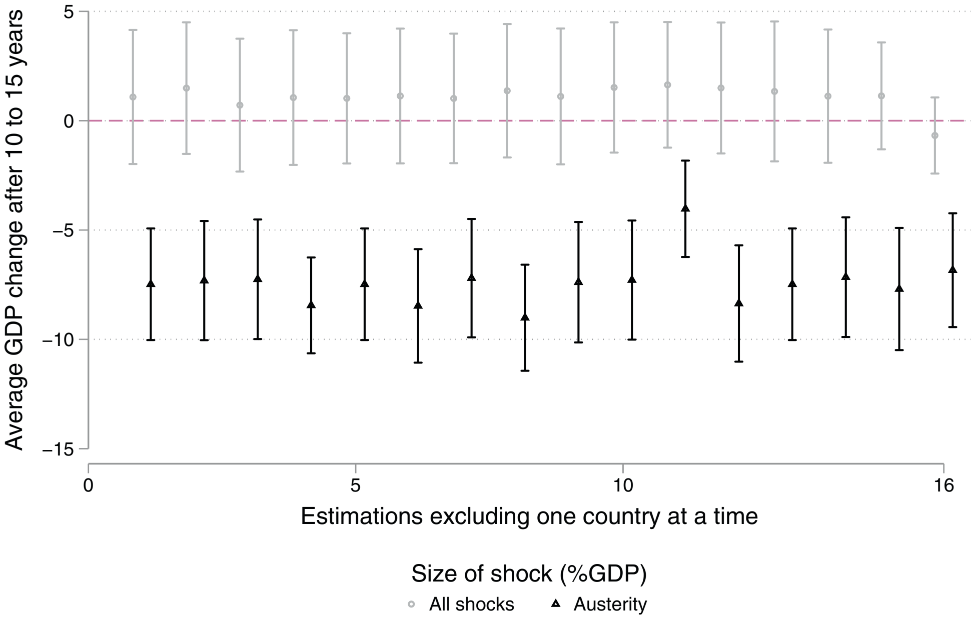

Figure 8. Robustness check – excluding countries. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

for the exclusion of any country. It is worth noting that the impact of austerity policies is particularly significant in Ireland, where strong contractionary fiscal measures were implemented from 2009 to 2012.

Initial examination of channels

A detailed examination of the channels through which these long-run effects operate is beyond the scope of this paper. Here, we perform a first approximation to check the effects on the three main aggregate inputs to the GDP: capital stock, labor and factor productivity.

We assume a Cobb–Douglas function:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta},$$

and use data from the Penn World Table (PWT) for the capital stock (K) and labor input (L), defined as the total number of hours worked. The Total Factor Productivity (A) is calculated as the residual using the production function above. The output elasticities of capital and labor, α and β, are given by the respective factor remuneration (also sourced from the PWT), which follows from the maximization under perfect competition.[24]

As previously discussed, one way in which shocks of different sizes can have heterogeneous proportional effects on the economy is through investment if, for instance, investment decisions are more sensitive to GDP growth after a given threshold. Despite

[24]For simplicity, one can think of the problem of a central planner: $Max\ A_{it} K_{it}^{\alpha} L_{it}^{\beta} - wL - rK$. The first-order conditions are: $\beta = wL/Y$ and $\alpha = rK/Y$.

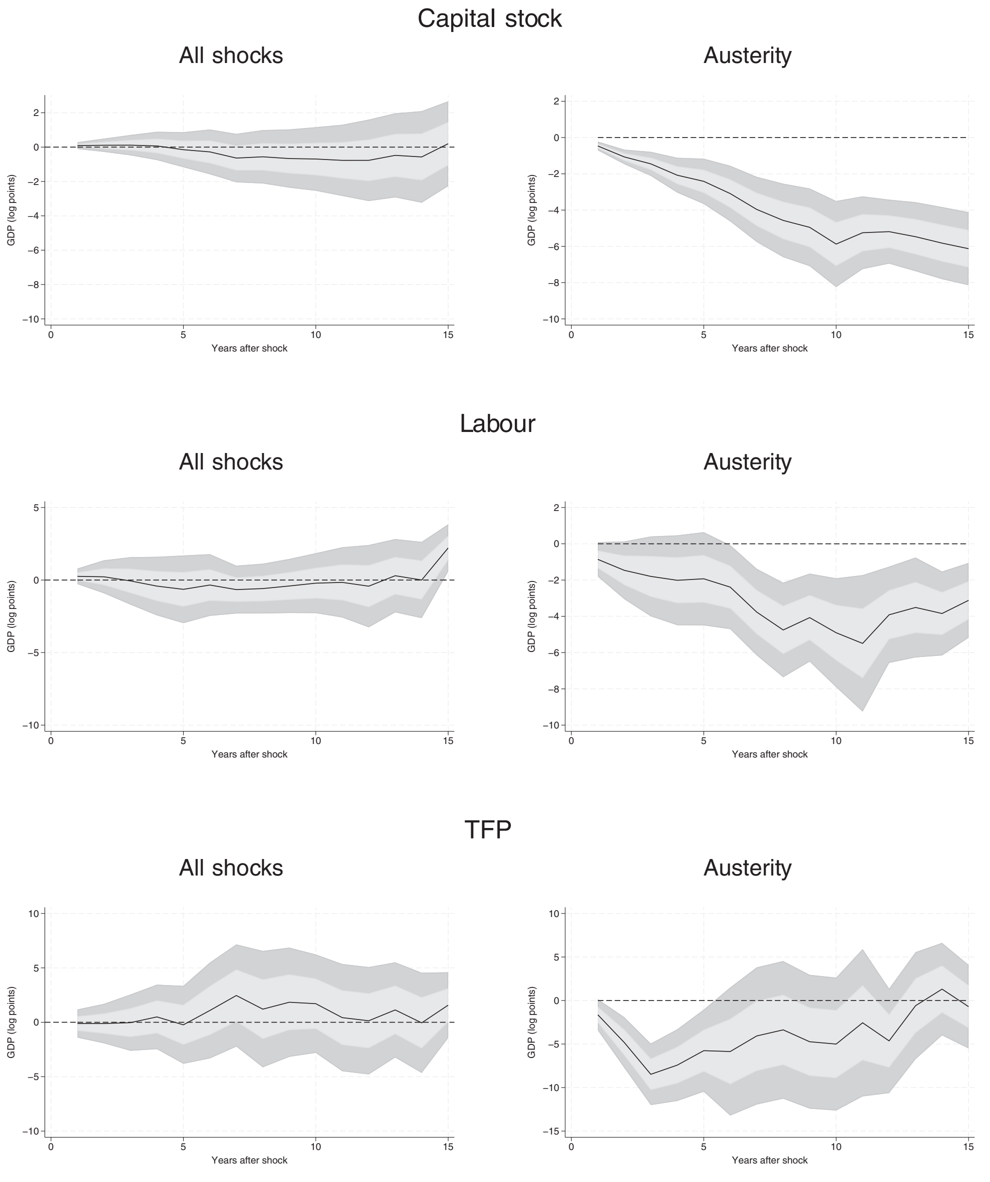

Figure 9. Effect by austerity shock by GDP component
Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by 2 SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP. The capital stock is sourced from the Penn World Table 10.0.

this potential non-linearity, the direction of the effect on the capital stock itself is not clear *a priori*, as government expenditure can crowd in or crowd out private investment depending on multiple factors, such as the type of fiscal shock and the effect on the interest rates, for instance (e.g. Antolin-Diaz and Surico, 2024; Bahal *et al.*, 2018). Figure 9 strongly suggests that austerity shocks are indeed associated with a consistent and statistically significant negative effect on the stock of capital, which could help explain the long-run effects on GDP. There is also a negative and significant effect of austerity on the labor used in the economy. Although weaker than the effect on the capital stock, the reduction in the number of employed people, even after 15 years of an austerity shock, helps us understand the aggregated effect on GDP. In both cases, it is interesting to note not only the effect of austerity shocks but also how different they are when we take into account all fiscal shocks. Finally, aggregate TFP is the factor that reacts the strongest immediately after the shock, contributing negatively to the GDP until at least the sixth year after the shock, but its effect converges to zero in the long run.

'Cleaner' controls – a local projections approach to DiD

An increasingly recognized problem in studies that resort to some form of differences-in-differences (DiD) estimation is the bias that emerges once one moves away from a '2X2' setup – that is, two periods (pre- and posttreatment) and two status (treated or never treated) (e.g., Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). In our case, one can illustrate an important potential bias by reminding that the regression that estimates the effect of an austerity shock in time t on output in time $t+k$ has as controls countries that also had shocks between $t+1$ and $t+k-1$. In situations in which the treatment effects are heterogeneous and dynamic, as in our case, the bias is clear: the observations used as controls are also under the influence of shocks.

There are different ways of trying to reduce this bias. The method suggested by Dube *et al.* (2022) seems particularly interesting and adequate for our purposes given the endogenous nature of the treatment time. In this section, we follow their approach by excluding from the control sample countries that were 'treated' between $t+1$ and $t+k$ when estimating the effect of treatment in t on output at $t+k$.[25] This is performed with our baseline setting (section III), that is, on top of performing propensity-score matching and controlling for future shocks of the treated countries.

Although this approach has the advantage of providing control units that are not under the influence of austerity, it comes with the relatively high cost of significantly decreasing the number of observations for each estimation. This might lead to a less smooth sequence of coefficients and a wider confidence interval. In our case, the smaller the threshold for the minimum shock size, the stricter the rule on controls will be.[26] We focus, therefore, only on the austerity shocks so that we can have an adequate number of observations. In Appendix C, we test a much less strict method to 'clear' the controls and our baseline results continue to hold.

Reassuringly, results are qualitatively the same as the baseline ones, with austerity shocks having long-run negative effects on GDP, the capital stock and labor input

[25]For instance, assume several countries have austerity shocks in 1990. To calculate the average effect of these episodes after 10 years, the control sample will consist only of countries that did not have an episode between 1990 and 2000. Similarly, to calculate the effects after five years, the control sample would consist of countries that did not experience an episode between 1990 and 1995.

[26]That is, for a smaller threshold, we have a larger number of shocks, and thus the number of countries that can be used as controls in a 15-year window is very reduced.

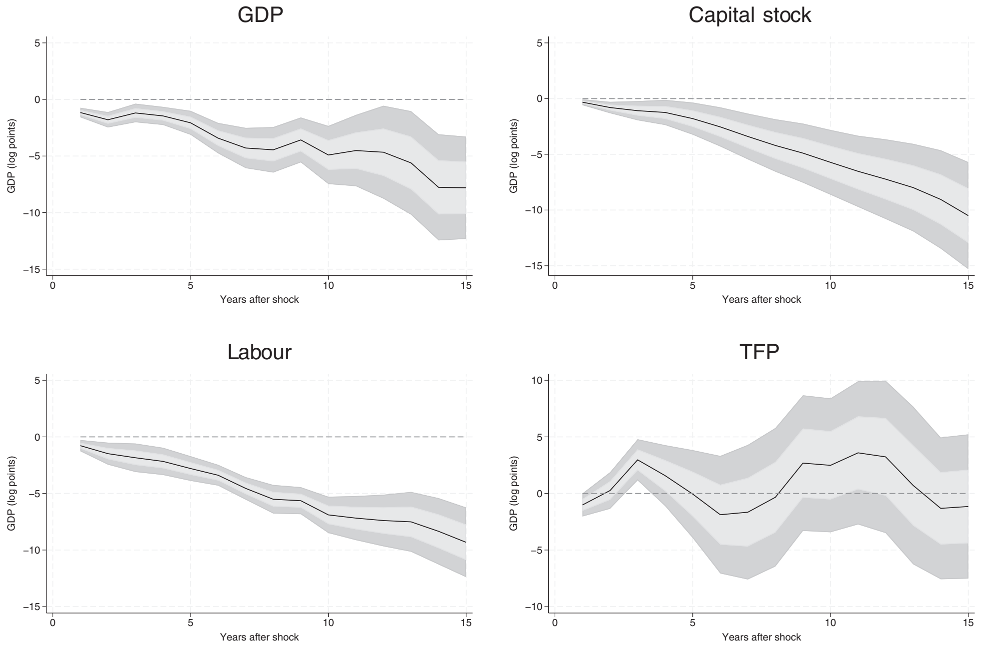

Figure 10. Cleaner controls – Austerity shocks
Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by 2 SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP.

(Figure 10). Moreover, the relatively small magnitude differences are in line with what would be expected: when considering only 'clean' controls, the effects of austerity are even stronger.

Estimation bias, lag length selection and penalized LPs

In our baseline estimation, we deal with two important sources of potential bias. One is that countries affected by austerity might have different characteristics than countries not affected; and the other is that within the 15-year horizon, a country might be affected by more than one austerity shock. To deal with the first, we reweight the sample using propensity score matching; to address the second, we control for future shocks.

In terms of the estimation method, however, another aspect has to be analyzed. As mentioned in section II, we prioritize the use of LPs as they tend to produce less biased estimations, although at the cost of larger variances (e.g. Li *et al.,* 2024). However, recent research has also indicated that lag-augmented LPs tend to be superior to LPs, particularly for long-run estimations (e.g. Montiel Olea and Plagborg-Møller, 2021; Antolin-Diaz and Surico, 2024), the basic idea being that lagged endogenous variables might be relevant to predict the treatment shock.

In our baseline estimations, we choose to use the simpler LP because (i) the AIPW model should account for part of this predictability, (ii) our dependent variable is in levels, which, as shown by Antolin-Diaz and Surico (2024), tends to reduce estimation biases

TABLE 4

Lag-augmented LPs

	Instantaneous (1 year)			*Long-run (15 years)*		
Number of lags	*0*	*4*	*10*	*0*	*4*	*10*
All shocks						
Effect	0.171	0.257	0.233	1.178	1.200	−0.205
Obs.	508	457	356	276	228	135
Austerity						
Effect	−0.151	−0.389***	−0.380***	−5.610**	−3.359**	−1.794***
Obs.	416	377	299	234	195	117

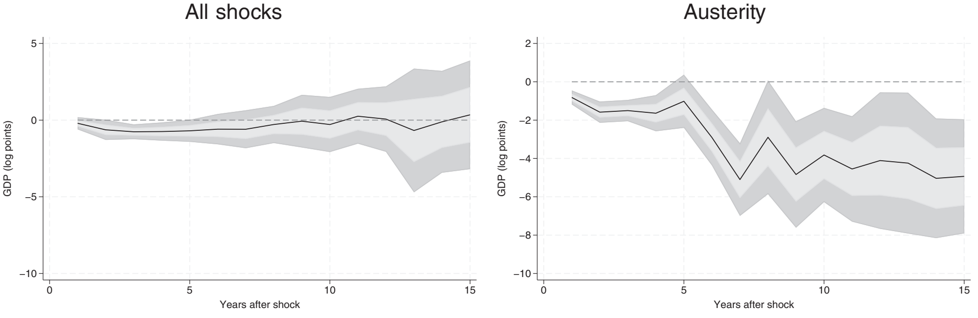

Figure 11. Penalized LPs

Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by 2 SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP.

(as compared to de-trended variables), but, importantly, and similarly to the discussion in section IV, (iii) the inclusion of many lags comes with a cost in terms of observations, which is especially relevant in a sample like ours, where each country is observed for only 36 years, and for long-run estimations.

As a robustness exercise, we test the effect of adding four and ten lags of the control variables in the LPs. As can be seen in Table 4, our baseline results, for all shocks and austerity, 1 year ahead and in the long-run, continue to hold, with austerity shocks generating statistically significant negative effects on GDP even after 15 years; the coefficient, however, tends to be smaller with the use of more lags. The table also indicates the problem with the reduction of the sample.

With respect to the trade-off between bias and variance mentioned a couple of times, Barnichon and Brownlees (2019) demonstrate that the use of penalized LPs can increase inference precision. As a robustness check, we estimate the long-run effects of the shocks using a penalized LP based on Friedman *et al.* (2010), which suggests a technique that incorporates elements of Lasso and Ridge methods.[27] As can be seen in Figure 11, the results are very similar to the baseline ones.

[27] In this estimation, instead of reweighting the sample, we add the variables used in the probit in the baseline method as controls in the main regression.

V. Conclusion

After a time of diminished interest in fiscal policy during the so-called Great Moderation in advanced economies, the past two decades saw an emerging interest in fiscal research, deriving from the challenges most economies faced since the Global Financial Crisis. Despite several efforts, which greatly improved our knowledge about the topic, a few important gaps persist. This paper aimed to address one in particular: the long-run effects of austerity policies; and, by doing so, shed some light on two other important questions: the heterogeneity of fiscal multipliers by shock size, and the persistent effects of demand shocks.

The idea that countries are still being affected by the most recent austerity wave that followed the financial crisis is widespread in public opinion. This impression might have encouraged the emergence of a literature that links austerity with several effects, including those that tend to have persistent impacts, from public health, to political instability and democracy erosion (e.g. Fetzer, 2019; Baccaro *et al.,* 2021; Ponticelli and Voth, 2020; Guriev and Papaioannou, 2022; Rajmil *et al.,* 2020). Regarding its economic impact, however, the evidence is limited to short-run effects.

Employing a method that 're-randomize' the allocation of austerity episodes in a local projections set up and accounting for the fact that multiple shocks occur in the time horizon of interest, our results indicate that austerity measures (defined as negative shocks larger than 3% of the GDP) have a detrimental effect on GDP of about 5.5% even after 15 years. This result is robust to extensions in the fiscal shocks used as controls, to different measures of GDP, to alternative narrative datasets, to the exclusion of individual shocks, and to the use of 'cleaner' controls. Moreover, there is robust evidence that austerity shocks have significant negative effects on capital stock and hours worked.

Finally, evidence that fiscal multipliers depend on the size of the shock is also presented: while the long-run multiplier for all shocks is 0.5, the one from austerity shocks is almost three times larger, 1.45.

This paper fills a relevant gap in the literature by: (i) examining the long-run effects of fiscal policy in general, employing techniques that are appropriate for such estimations; and (ii) allowing different effects for different shock sizes, both in proportional terms and related to its persistence over time. This last point, besides contributing to the emerging literature on heterogenous fiscal multipliers, is particularly relevant as the term 'austerity' is of public interest and, thus, it is important that economists engage in the broader conversation with a similar understanding of the term: contractionary fiscal policy of significant size. Arguing, *a priori*, that standard fiscal multipliers are sufficient to assess the impact of austerity episodes is misleading, does not contribute to our understanding of the topic and is not very useful for policy orientation. Finally, when it comes to the time horizon of the estimation, our study contributes to the growing literature on the persistent effects of demand shocks by being the first to analyse the long-run impact of narrative fiscal shocks. In this context, our estimations present additional evidence that demand shocks may have significant long-run effects.

Final Manuscript Received: April 2024

References

Acemoglu, D., Naidu, S., Restrepo, P. and Robinson, J. A. (2019). 'Democracy does cause growth', *Journal of Political Economy*, Vol. 127, pp. 47–100.

Alesina, A. and Ardagna, S. (2010). 'Large changes in fiscal policy: taxes versus spending', *Tax policy and the economy*, Vol. 24, pp. 35–68.

Alesina, A., Azzalini, G., Favero, C., Giavazzi, F. and Miano, A. (2018). 'Is it the ''how'' or the ''when'' that matters in fiscal adjustments?', *IMF Economic Review*, Vol. 66, pp. 144–188.

Alesina, A., Favero, C. and Giavazzi, F. (2019a). *Austerity: When it Works and When it Doesn't*, New Jersey: Princeton University Press.

Alesina, A., Favero, C. and Giavazzi, F. (2019b). 'Effects of austerity: expenditure-and tax-based approaches', *Journal of Economic Perspectives*, Vol. 33, pp. 141–162.

Antolin-Diaz, J. and Surico, P. (2024). 'DP17433 The Long-Run Effects of Government Spending', CEPR Discussion Paper No. 17433. CEPR Press, Paris & London.

Auerbach, A. J. and Gorodnichenko, Y. (2012). 'Measuring the output responses to fiscal policy', *American Economic Journal: Economic Policy*, Vol. 4, pp. 1–27.

Auerbach, A. J. and Gorodnichenko, Y. (2013). 'Fiscal Policy after the Financial Crisis', edited by Alberto Alesina and Francesco Giavazzi, Chicago: University of Chicago Press, pp. 63–102. https://doi.org/10.7208/9780226018584-004.

Baccaro, L., Bremer, B. and Neimanns, E. (2021). 'Till austerity do us part? a survey experiment on support for the euro in Italy', *European Union Politics*, Vol. 22, pp. 401–423.

Bahal, G., Raissi, M. and Tulin, V. (2018). 'Crowding-out or crowding-in? public and private investment in India', *World Development*, Vol. 109, pp. 323–333.

Ball, L. M., Furceri, D., Leigh, M. D. and Loungani, M. P. (2013). *The Distributional Effects of Fiscal Consolidation*, International Monetary Fund. No. 2013/151. International Monetary Fund, 2013.

Barnichon, R. and Brownlees, C. (2019). 'Impulse response estimation by smooth local projections', *Review of Economics and Statistics*, Vol. 101, pp. 522–530.

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

Baum, M. A., Poplawski-Ribeiro, M. M. and Weber, A. (2012). *Fiscal Multipliers and the State of the Economy*, International Monetary Fund. IMF Working Papers, Vol. 12, no. 286, p. 1.

Ben Zeev, N., Ramey, V. A. and Zubairy, S. (2023). 'Do government spending multipliers depend on the sign of the shock?', in *AEA Papers and Proceedings* Vol. 113, American Economic Association, Nashville, TN, pp. 382–387.

Blanchard, O. and Perotti, R. (2002). 'An empirical characterization of the dynamic effects of changes in government spending and taxes on output', *The Quarterly Journal of Economics*, Vol. 117, pp. 1329–1368.

Blanchard, O. and Quah, D. (1989). 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review*, Vol. 79, pp. 655–673.

Blanchard, O., Cerutti, E. and Summers, L. (2015). *Inflation and Activity–Two Explorations and Their Monetary Policy Implications*, Technical Report, National Bureau of Economic Research. Working Paper Series, number 21726.

Callaway, B. and Sant'Anna, P. H. (2021). 'Difference-in-differences with multiple time periods', *Journal of Econometrics*, Vol. 225, pp. 200–230.

Cerra, V. and Saxena, S. C. (2008). 'Growth dynamics: the myth of economic recovery', *American Economic Review*, Vol. 98, pp. 439–457.

David, A. and Leigh, D. (2018). *A New Action-based Dataset of Fiscal Consolidation in Latin America and the Caribbean*, International Monetary Fund. IMF Working Paper No. 18/94, Available at SSRN: https://ssrn.com/abstract=3183301.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review*, Vol. 110, pp. 2964–2996.

Devries, P., Guajardo, J., Leigh, D. and Pescatori, A. (2011). *A New Action-Based Dataset of Fiscal Consolidation*, International Monetary Fund. IMF Working Paper No. 11/128, Available at SSRN: https://ssrn.com/abstract=1861798.

Dube, A., Girardi, D., Jorda, O. and Taylor, A. (2022). *A Local Projections Approach to Difference-in-Differences Event Studies*, NBER Working Paper No. w31184, Available at SSRN: https://ssrn.com/abstract=4433863.

Fatás, A. and Summers, L. H. (2018). 'The permanent effects of fiscal consolidations', *Journal of International Economics*, Vol. 112, pp. 238–250.

Fehr, E. and Tyran, J.-R. (2001). 'Does money illusion matter?', *American Economic Review*, Vol. 91, pp. 1239–1262.

Fetzer, T. (2019). 'Did austerity cause brexit?', *American Economic Review*, Vol. 109, pp. 3849–3886.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). 'Regularization paths for generalized linear models via coordinate descent', *Journal of statistical software*, Vol. 33, p. 1.

Gechert, S., Horn, G. and Paetz, C. (2019). 'Long-term effects of fiscal stimulus and austerity in europe', *Oxford Bulletin of Economics and Statistics*, Vol. 81, pp. 647–666.

Ghassibe, M. and Zanetti, F. (2022). 'State dependence of fiscal multipliers: the source of fluctuations matters', *Journal of Monetary Economics*, Vol. 132, pp. 1–23.

Giavazzi, F. and Pagano, M. (1990). 'Can severe fiscal contractions be expansionary? tales of two small european countries', *NBER Macroeconomics Annual*, Vol. 5, pp. 75–111.

Goodman-Bacon, A. (2021). 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics*, Vol. 225, pp. 254–277.

Greenwald, B. C., Stiglitz, J. E., Hall, R. E. and Fischer, S. (1988). 'Examining alternative macroeconomic theories', *Brookings Papers on Economic Activity*, Vol. 1988, pp. 207–270.

Gross, D. P. and Sampat, B. N. (2023). 'America, jump-started: World war II R&D and the takeoff of the us innovation system', *American Economic Review*, Vol. 113, pp. 3323–3356.

Guajardo, J., Leigh, D. and Pescatori, A. (2014). 'Expansionary austerity? international evidence', *Journal of the European Economic Association*, Vol. 12, pp. 949–968.

Guriev, S. and Papaioannou, E. (2022). 'The political economy of populism', *Journal of Economic Literature*, Vol. 60, pp. 753–832.

Haltmaier, J. (2013). Do Recessions Affect Potential Output? FRB International Finance Discussion Paper No. 1066.

Hamilton, J. D. (2018). 'Why you should never use the hodrick-prescott filter', *Review of Economics and Statistics*, Vol. 100, pp. 831–843.

Ilzetzki, E., Mendoza, E. G. and Végh, C. A. (2013). 'How big (small?) are fiscal multipliers?', *Journal of Monetary Economics*, Vol. 60, pp. 239–254.

Iwata, Y. and IIboshi, H. (2023). 'The nexus between public debt and the government spending multiplier: fiscal adjustments matter', *Oxford Bulletin of Economics and Statistics*, Vol. 85, pp. 830–858.

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

Jordà, Ò. and Taylor, A. M. (2016). 'The time for austerity: estimating the average treatment effect of fiscal policy', *The Economic Journal*, Vol. 126, pp. 219–255.

Jordà, Ò., Singh, S. R. and Taylor, A. M. (2020). *The Long-Run Effects of Monetary Policy*, Technical Report, National Bureau of Economic Research. Working Paper Series, number 26666.

Kantor, S. and Whalley, A. T. (2023). *Moonshot: Public R&D and Growth*, Technical Report, National Bureau of Economic Research. Working Paper Series, number 31471.

Li, Z. and Koustas, D. (2019). 'The long-run effects of government spending on structural change: evidence from second world war defense contracts', *Economics Letters*, Vol. 178, pp. 66–69.

Li, D., Plagborg-Møller, M. and Wolf, C. K. (2024). 'Local projections vs. vars: lessons from thousands of dgps', *Journal of Econometrics*, 105722.

Lucas, R. E., Jr. (2003). 'Macroeconomic priorities', *American Economic Review*, Vol. 93, pp. 1–14.

Lunceford, J. K. and Davidian, M. (2004). 'Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study', *Statistics in medicine*, Vol. 23, pp. 2937–2960.

Martin, R., Munyan, T. and Wilson, B. A. (2015). *Potential Output and Recessions: Are We Fooling Ourselves?* FRB International Finance Discussion Paper No. 1145.

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

Mountford, A. and Uhlig, H. (2009). 'What are the effects of fiscal policy shocks?', *Journal of Applied Econometrics*, Vol. 24, pp. 960–992.
Ponticelli, J. and Voth, H.-J. (2020). 'Austerity and anarchy: budget cuts and social unrest in europe, 1919–2008', *Journal of Comparative Economics*, Vol. 48, pp. 1–19.
Rajmil, L., Hjern, A., Spencer, N., Taylor-Robinson, D., Gunnlaugsson, G. and Raat, H. (2020). 'Austerity policy and child health in european countries: a systematic literature review', *BMC Public Health*, Vol. 20, pp. 1–9.
Ramey, V. (2011). 'Identifying government spending shocks: it's all in the timing', *The Quarterly Journal of Economics*, Vol. 126, pp. 1–50.
Ramey, V. (2019). 'Ten years after the financial crisis: what have we learned from the renaissance in fiscal research?', *Journal of Economic Perspectives*, Vol. 33, pp. 89–114.
Ramey, V. and Shapiro, M. (1998). 'Costly capital reallocation and the effects of government spending', in *Carnegie-Rochester Conference Series on Public Policy* Vol. 48, Elsevier, North-Holland, pp. 145–194.
Ramey, V. and Zubairy, S. (2018). 'Government spending multipliers in good times and in bad: evidence from us historical data', *Journal of Political Economy*, Vol. 126, pp. 850–901.
Riera-Crichton, D., Vegh, C. A. and Vuletin, G. (2016). 'Tax multipliers: pitfalls in measurement and identification', *Journal of Monetary Economics*, Vol. 79, pp. 30–48.
Romer, C. and Romer, D. (2010). 'The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks', *American Economic Review*, Vol. 100, pp. 763–801.
Solow, R. M. (1956). 'A contribution to the theory of economic growth', *The Quarterly Journal of Economics*, Vol. 70, pp. 65–94.
Stenner, N. (2022). 'The asymmetric effects of monetary policy: evidence from the United Kingdom', *Oxford Bulletin of Economics and Statistics*, Vol. 84, pp. 516–543.
Tenreyro, S. and Thwaites, G. (2016). 'Pushing on a string: US monetary policy is less powerful in recessions', *American Economic Journal: Macroeconomics*, Vol. 8, pp. 43–74.
Teulings, C. and Zubanov, N. (2014). 'Is economic recovery a myth? robust estimation of impulse responses', *Journal of Applied Econometrics*, Vol. 29, pp. 497–514.
Yellen, J. (2016). *Macroeconomic Research After the Crisis: a speech at ''The Elusive 'Great' Recovery: Causes and Implications for Future Business Cycle Dynamics'' 60th annual economic conference sponsored by the Federal Reserve Bank of Boston, Boston, Massachusetts, October 14, 2016*. No. 915, Board of Governors of the Federal Reserve System (US).

Appendix A: Exogeneity

Using the original dataset, we can see that the probability of treatment is not randomly assigned; for instance, units that were treated in the previous year have higher probabilities of treatment. However, after rebalancing the sample using the method described in the main text, the probability of treatment cannot be anticipated by any of the variables of interest.

An additional exercise to analyse the exogeneity of the austerity shock is the so-called Durbin-Wu-Hausman Test (Figure A1). Initially, we perform a regression where the austerity shock is the dependent variable and the variables used in the probit regression of the baseline method are the predictors, to capture the portion of shock explained by these variables (Table A1). The residuals from this regression represent the part of austerity shock not explained by the instruments.

Next, we include these residuals in the main regression model along with the austerity shock and the other control variables and test whether these residuals are statistically significant. If they are, it indicates that the austerity shocks is endogenous, meaning it is correlated with the error term in the main regression, and hence needs to be treated differently to avoid biased results. If not, the austerity shock can be considered exogenous, implying it does not suffer from endogeneity and the initial regression estimates are valid.

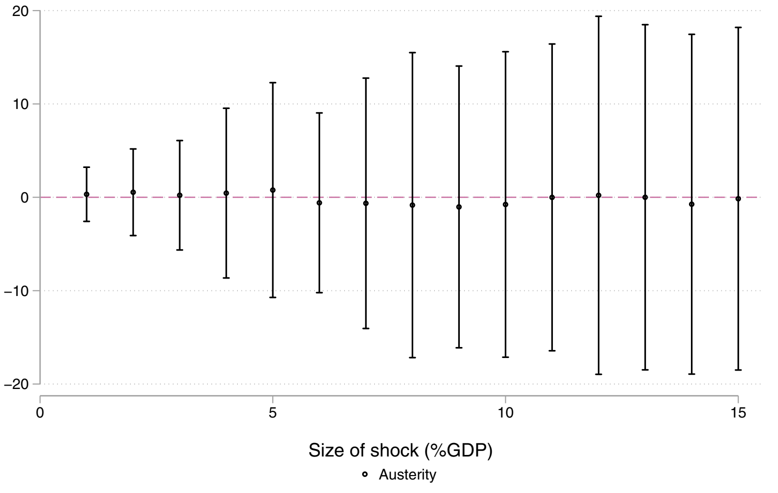

Figure A1. Durbin-Wu-Hausman Test. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Note: The black dot is the residual's coefficient, and the lines give the 90% confidence interval

TABLE A1
Probit models with and without inverse probability weights

	Before	*After*
debt_all	0.287	−0.784
	(0.870)	(0.884)
hply	0.143*	−0.023
	(0.077)	(0.065)
dly	−0.070	−0.054
	(0.069)	(0.066)
ldly	−0.156***	−0.044
	(0.058)	(0.064)
treatment	1.212***	−0.049
	(0.329)	(0.508)
dlcpi	−0.073	−0.025
	(0.060)	(0.067)
dlriy	0.005	−0.014
	(0.034)	(0.033)
stir	−0.219**	-0.090
	(0.087)	(0.087)
ltrate	0.283***	0.078
	(0.100)	(0.107)
cay_all	0.002	0.000
	(0.003)	(0.001)
drprv_all	−0.023	0.002
	(0.023)	(0.018)
Observations	442	441

Note: SEs in parentheses.
* $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

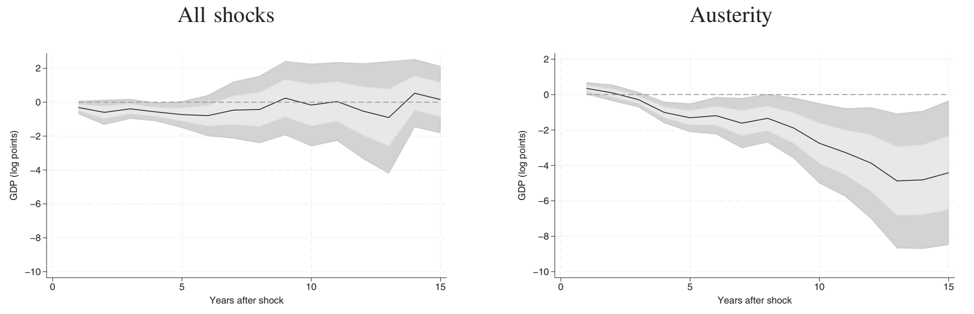

Figure B1. Effect using unexplained residual from narrative shock
Note: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by 2 SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval)

The results indicate that we cannot reject the hypothesis that the residuals are equal to zero. That is, that the narrative fiscal shocks, using the reweighted sample and the additional controls in the outcome regression, is exogenous.

Appendix B: Using residuals as shocks

In the first stage, we regress the narrative fiscal shock on the other independent variables. We then use the residual of this regression as the treatment in the regression that follows the same structure as our baseline one. The number of total shocks is reduced to 203, and austerity shocks to 53. Now, austerity is defined as residuals larger than 0.5% of GDP.

As shown in Figure B1, the outcomes closely resemble our baseline results, reinforcing our confidence that the effects of the fiscal shocks are not driven by other control variables.

Appendix C: Comparison with non-austerity periods

The baseline results reflect the average treatment effect compared to those not treated. That is, in the case of all shocks, it is compared to those that do not have any shock in the same year, and in the case of austerity (Figure C1), those that not have an austerity shock in the same year.

In this exercise, we included only treated units and those that have no fiscal constraints at all (i.e., of any size) in the regression. When dealing with 'all shocks', this represents the same specification, as each unit is either having some fiscal constraint or not. For the austerity estimation, however, this specification represents a change, as now units that have a fiscal constraint that are not large enough to be considered an austerity shock are also dropped out from the sample. As can be seen, results are virtually the same.

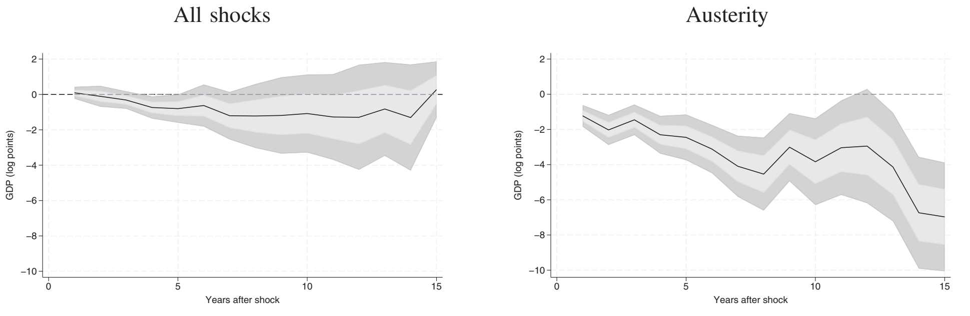

Figure C1. Effect of Austerity – by size of the shock
Notes: The black lines indicate estimated coefficients. The darker grey area is the confidence interval given by 2 SDs (approximately 95% confidence interval), while the light grey area is the confidence interval given by 1 SD (approximately 68% confidence interval). Austerity is defined by shocks larger than 3% of the GDP.