Long-run effects of austerity

Long-run effects of austerity*

Guilherme Klein Martins†

November 30, 2022

Abstract

This paper provides evidence that austerity shocks have long-run negative effects on GDP. Besides addressing the important gap in the growing fiscal research regarding the short time horizon of the estimations, this paper analyzes two other important assumptions made in the literature regarding the (i) symmetry of episodes of fiscal expansion and contraction and (ii) uniformity of fiscal multipliers for different sizes of shocks. We use narrative fiscal shocks and propensity score reweighting in a local projections setup to account for the potential endogeneity of austerity policies and the non-linearity of its effects over time. The estimation is also adapted to eliminate the bias that emerges when multiple shocks might occur within the time horizon of interest. Our baseline results show that contractionary fiscal shocks larger than 1.5% of GDP generate a negative effect of more than 3% on GDP even after 15 years. The drop in GDP reaches 5.5% for fiscal contractions larger than 3%. Evidence is also found linking austerity with smaller capital stock in the long-run. The results are robust to different fiscal shocks datasets, the exclusion of particular countries and shocks, alternative estimation methods, and the use of cleaner controls. Besides understanding the consequences of this particular policy, the results contribute to the broader discussion on the long-run effects of demand by suggesting that such shocks might permanently affect the economy.

JEL codes: C54, E62, H5, H2

Key words: austerity, long-run, fiscal policy, multipliers

*I thank Daniele Girardi and Laura Carvalho for critical suggestions to this paper. I also thank Peter Skott, Arslan Razmi, and Christian Rojas for helpful comments on earlier versions. All remaining errors are mine.

†Department of Economics, University of Massachusetts Amherst, gmartins@umass.edu.

(...) 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)

1 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 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 labor 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.

However, it is clear that to evaluate the success of the austerity strategy it is not sufficient to compare averages. Ideally, 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.

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

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

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 V. 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[3] of (ii) fiscal shocks in general.

There might be different reasons for the shorter-run focus. V. Ramey (ibid.) 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, Cerutti, 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 analyze 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 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, Favero, et al. (2019a): "[t]he term "austerity" indicates a policy of *sizeable reduction* of government deficits and stabilization of

[3]This might be a too general statement, but exceptions seem to be extremely rare indeed. One that could be cited is Fatás and Summers (2018), that look exclusively to consolidations that took place in 2010-2011 and whose estimations are completely based on forecasts, both for GDP (up to 2021) and for the structural balance. Although informative, these estimations seem to be significantly less robust than those that take into account a much larger number of shocks, use methods to achieve shocks that are as exogenous as, arguably, possible, and that do not rely on forecast errors, for instance.

government debt achieved by means of spending cuts or tax increases, or both." (p.1, italics added). The literature, however, with very few exceptions,[4] ignores this definition in two important ways by assuming that (i) fiscal contractions and expansions are symmetrical, and (ii) that the effects are linear on the size of the shock.[5]

There are multiple theoretical reasons why these two aspects might be relevant. The recognition that positive and negative demand shocks tend to have asymmetrical effects is not new (e.g., De Long et al. (1988), Cover (1992)). There are different channels through which this could operate. The economy can have multiple equilibriums[6], with positive and negative shocks pushing the economy to different 'steady-states'. Another channel is more explicit in efficiency wages models (e.g., Summers (1988)), in which workers quickly adjust their wage expectations upwards after a positive shock but do not do so following a negative one; the validity of this channel is reinforced by research on behavioral economics, for instance regarding self-serving biases (Babcock and Loewenstein (1997)) and money illusion (Fehr and Tyran (2001)). The effects might be asymmetrical also due to different reactions of the financial market: as in Greenwald et al. (1988), banks can either remain healthy - the probable outcome of a positive shock -, or fail - a possible result of negative shocks. While the first situation might not generate permanent effects, the second tends to do so.

Regarding heterogeneous effects by the shock size, most of these reasons can also be important: shifts between equilibriums might depend on the size of the initial departure from the former equilibrium; the cognitive costs related to operating with nominal or real values are non-linear[7]; financial institutions are resilient to relatively small negative shocks. 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 direction and the size of the fiscal shock is important

[4] Alesina and Ardagna (2010) is an important exception, as the authors calculate separatedly the effects of expansions and contractions using the CAPB method.

[5] As will be resumed in section 2, in some sense the size of the shock is relevant for an important strand of the literature, as in Alesina, Favero, 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.

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

[7] Money illusion is more probable for relatively low levels of inflation.

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. A more detailed description of such theoretical reasons is beyond the scope of the paper. The discussion of how the empirical literature deals with these dimensions is resumed in section 2.

This paper aims to fill this important gap 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 1.5% of GDP in the baseline case and 1% of GDP in robustness exercises) generate a significant and persistent reduction in GDP even after 15 years; this result is robust to the use of alternative datasets (both of extended GDP and austerity shocks), the exclusion of countries and episodes, and the implementation of different estimation methods. There is also evidence that short- and long-run multipliers are different for relatively small and large shocks. We also find indications that spending cuts generate larger negative effects on GDP, and that austerity shocks are associated with lower capital stock.

Besides this introduction, the paper has three other parts. In section 2, 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 3 explains the method and data in more detail and presents our baseline estimations. It is followed by section 4, in which a series of robustness checks and extensions are performed. Section 5 concludes the paper.

2 Fiscal Research

To help organize the literature, one can point, in line with V. Ramey (2019), to the three main methods used in empirical fiscal research: i) aggregate country-level time series or panel estimates, ii) estimated or calibrated New Keynesian DSGE models, and iii) 'natural experiment approaches' that use, for instance, variations in sub-national units for identification.[8] Each of those has its weaknesses: time series methods require exogenous variation in policy, which sometimes forces the use of inadequate instruments;[9] estimations based on DSGE models, on

[8] Using war-induced government spending for identification can also be fitted in this category; however, this method does not apply well to other countries with lower defense spending or for those that the fluctuations are associated with conflict within the country.

[9] That are either exogenous but not very relevant (have a low correlation to the fiscal variable) or relevant but not exogenous or unanticipated. An example of the first type is military news, which are weak instruments after 1954 (V. A. Ramey (2011)); an example of the latter is the one-step ahead forecast error of government spending, used in Blanchard and Perotti (2002).

their part, rely on strong assumptions about the generating process of unobserved shocks and the theoretical structure. Moreover, subnational analyses do not lead directly to macroeconomic estimations, also requiring some theoretical model to do this passage.

All considered, the literature, following efforts to improve the main weakness of the method and capture shocks that are as exogenous as possible, has been converging to the use of country-level data of exogenous policy changes. 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. C. Romer and D. Romer (2010) point out that CAPB is affected by nonpolicy changes that might be correlated with other elements affecting output[10]. 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)). Caveats can also be made on the subjectivity of the method to extract the economic cycle out of data (and how estimations tend to be sensitive to this choice), as well as the usual assumption of a constant elasticity of expenditure to the economic cycle (e.g., C. Romer and D. Romer (2010); Agnello and Sousa (2014)).

An alternative to CAPB[11] 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 the changes motivated by the goal of increasing long-run growth or reducing the budget deficit.[12]

[10] 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.

[11] 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.

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

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: a higher weight is given to countries that, although having a higher probability of having a shock, do not have one. At the current stage of the literature, this combination of narrative fiscal shocks and propensity weighting seems to be the best strategy to analyze fiscal shocks, and is, therefore, the one employed in this paper. More details of the method will be presented in section 3.

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.[13] 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. (2022); Jordà, Singh, et al. (2020)).[14]

After this brief overview of the state of the literature, we can return to the observation by V. Ramey (2019), mentioned in section 1, 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, Singh, 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 thirty-years 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

[13] 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.

[14] V. A. Ramey and Zubairy (2018) use the paper of Auerbach and Gorodnichenko (2012b) 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 (2012b), ones 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 (2012a) 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.

shocks. Additionally, in this paper, we adapt our estimations to account for a potential bias that emerges when multiple shocks occur within the forecasted horizon. Following the suggestion by Teulings and Zubanov (2014), this consists of controlling for a flexible number of treatment leads in the local projections regressions. The econometric strategy is explained in more detail in section 3.

As also indicated in section 1, 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.[15] In recent years a number of papers resumed the discussion about the long-term effects of 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, Cerutti, et al. (2015) over a maximum horizon of 10 years are some examples). However, there are no such estimations for fiscal shocks. Therefore, by estimating the long-run effects of austerity, this paper also contributes to the broader debate on the persistent effects of demand shocks.

Table 1 lists some of the most influential papers in the fiscal research literature. The literature is vast, and the list is produced to include papers closer to ours in estimating shocks using country-level data, but also to illustrate the diversity of empirical methods used. The most obvious difference between our estimation and the literature is, as addressed at length,

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

the maximum time horizon.[16]

Table 1: Selected studies of the effect of fiscal shocks on GDP

Authors	Data	Identification	Method	Max. Horizon
Alesina, Favero, et al. (2019b)	16 OECD Countries	Narrative	VAR	Five years
	1978-2014			
Jordà and Taylor (2016)	17 OECD Countries	Narrative	LP (AIPW)	Five years
	1978-2009			
Riera-Crichton et al. (2016)	15 OECD Countries	Narrative	LP	One year
	1980-2009	(VAT changes)		
Guajardo et al. (2014)	17 OECD Countries	CAPB	2SLS and VAR	Five years
	1978-2009	inst. by narrative		
Ilzetzki et al. (2013)	44 countries	CAPB	VAR	Five years
	1960-2007	(Expenditure)		
Baum et al. (2012)	6 OECD Countries	CAPB	TVAR	Three years
	1965-2011			
Auerbach and Gorodnichenko (2012b)	US	CAPB	STVAR	Five years
	1966-2009	inst. by forecast		
C. Romer and D. Romer (2010)	US	Narrative (Tax)	OLS and VAR	Five years
	1947-2007			
Mountford and Uhlig (2009)	US	Sign restriction	VAR	Six years
	1955-2000			

Note: AIPW: Augmented Inverse Propensity Weighted Estimator; TVAR: Threshold Vector Autoregression. STVAR is an extension of smooth transition autoregressive (STAR) models.

However, there is another important element that is common in these works and, as mentioned in section 1, is explored in this paper: the assumption of linearity of the effect of fiscal change. This assumption appears in two forms: that positive and negative shocks are taken to be symmetrical, and that shocks of different sizes have the same proportional effects. The symmetry assumption is explicit when the estimated effect of a fiscal shock is the average (weighted by the number of respective shocks) of the effects of positive and (inverted) negative shocks. Once one considers the theoretical reasons why the effects might not be symmetrical, such as the ones presented in section 1, it is clear how the assumption can be misleading. Assume, for instance, that in a sample there is the same number of fiscal expansions and contractions, and, to simplify, that the size of all shocks is 1% of GDP in absolute terms. Assume, finally, that the average effect of expansions is to increase GDP by 10% and contractions do not change GDP. In this case, grouping all estimations, we would get the result that an increase (decrease) in the

[16]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.

fiscal variable of 1% of GDP will increase (decrease) GDP by 5%. This, of course, is correct as an average effect. However, it obscures essential differences between the two types of policies.

The assumption that shocks of different sizes have proportional effects 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 (fiscal shock or without fiscal shock), as in Jordà and Taylor (2016). However, even in estimations with a 'continuous' treatment, that is, the size of the shock as the independent variable[17], for instance, a limitation persists. First, because 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 even for the most common narrative fiscal shocks datasets.

Secondly, because in the particular discussions about austerity measures, to which many of the papers listed participate, considering shocks of all sizes and assuming they have the same elasticities is misleading. As indicated in the introduction (section 1), the term 'austerity' carries a more or less specific meaning among economists and the general public, that of a *significant reduction* in government primary balance, and, not rarely, a more specific understanding of a reduction in public spending.[18] It must be acknowledged that the literature, by analyzing differences in the shocks led by taxes or spending changes, advanced significantly in understanding this latter aspect of what is sometimes taken to be austerity shocks in the short-run. The broader aspect related to the size of the shock, however, has not been explored yet.[19]

For these reasons, we believe that estimating separately the effects of (i) only contractionary shocks, and (ii) by different size brackets, is also an important contribution of this paper to the

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

[18]Examples are abundant. In this article, for instance, it is suggested that tax increases would be required to end austerity (that is, the reduction in public spending): https://www.theguardian.com/politics/2018/oct/04/is-austerity-really-over-theresa-mays-promise-lacks-key-details. In this New York Times article, austerity is defined as "a campaign of budget cutting" (https://www.nytimes.com/2019/02/24/world/europe/britain-austerity-may-budget.html). In this UN report, austerity is also associated with spending cuts: "austerity policies(...) eliminated many social services, reduced policing services to skeletal proportions, closed libraries in record numbers, shrunk community and youth centres, and sold off public spaces and buildings including parks and recreation centres" (https://www.ohchr.org/en/documents/country-reports/ahrc4139add1-visit-united-kingdom-great-britain-and-northern-ireland).

[19]It must be noted also that some earlier works that used the CAPB method were closer to our claim that the shock must be large enough to be considered an austerity shock. In Alesina and Ardagna (2010), for instance, it was considered a shock if the CAPB changed by more than 1.5% of GDP. The goal of the threshold, however, was to be sure one was capturing a shock and not to focus on large ones - it is relevant to note that the size of the shocks captured by the CAPB method are significantly larger: 2.4% of GDP in Alesina and Ardagna (ibid.) compared to 0.9% in Devries et al. (2011).

literature.

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, with a larger multiplier for tax changes than spending (V. Ramey (2019)). An important exception is a paper by Alesina and Ardagna (2010), which found that negative fiscal shocks had a positive effect on output in a three-year horizon, sparking an intense discussion around the "expansionary austerity" hypothesis. 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 labor market. Expenditure cuts (in government jobs and wages, for instance) would worsen workers' fallback position, decreasing wages in the private sector, allegedly increasing profits, investment and competitiveness. Increases in taxes, on the other side, would tend to increase the pretax real wage, squeezing profits, investment and competitiveness.

However interesting these theoretical channels might be, most of the papers that followed pointed in the opposite direction. Let us mention two that are closely related to ours. Jordà and Taylor (2016) first replicate, using LP, the results of Alesina and Ardagna (2010), but, the authors show that this result is driven entirely by the effects of contractionary policies during booms. The next step given by Jordà and Taylor (2016), and already mentioned, is to show that narrative episodes are not good instruments as they are also endogenous.[20] Given this, the statistical design proposed by Jordà and Taylor (ibid.) is the following: i) use the consolidation episodes identified in the IMF narrative dataset as the maximum subset of episodes (a 'pseudo-IV' step); ii) add the covariates that can predict the fiscal shock or influence output as controls; and iii) use inverse propensity score weighting to re-randomize the allocation of the consolidation episodes. With this setup, similar to the one employed in this paper, the authors find that

[20] First, they indicate that for a number of variables (Public debt to GDP ratio, deviation of log output from trend, output growth rate, and lagged value of treatment), the means are statistically different for 'treated' and control groups, indicating that the distribution of treatment is significantly different than an ideal randomised controlled trial. The authors also find that other variables, usually omitted in the regressions that try to identify the causal effect of fiscal shocks on output, are significant in explaining GDP fluctuations in a regression that also contains fiscal shocks (CAPB and its instrumentalized versions) as an independent variable. Finally, using different binary classification models, they show that the occurrence of fiscal consolidations (as indicated by the IMF narrative dataset) can be predicted by a number of variables (public debt to GDP ratio; the output gap; GDP growth; and fiscal consolidation itself).

consolidation episodes are associated with lower GDP within a five-year horizon.

Alesina, Favero, et al. (2019b) present the analysis[21] of austerity plans in 16 OECD countries from the 1970s to 2014 using the narrative approach by extending the dataset elaborated by Devries et al. (2011). They use panel vector autoregression approach to analyze the effect of such plans on a 4-year time horizon. They argue that while austerity based on spending cuts generates minor negative effects and only in the first year, plans based on tax increases reduce GDP by about 2% after four years.

Our paper can be placed within this large and emerging fiscal research literature. There are a number of gaps and issues, however, that this work aims at addressing. 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. Secondly, this paper does not assume, as the majority of the literature, that positive and negative shocks are symmetrical. Finally, 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 because the shocks are "too common" in the datasets, weakening identification.

3 Estimations

3.1 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, Azzalini, 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 United Kingdom, and the United States) from 1978 to 2014. The dataset by Alesina, Azzalini, et al. (ibid.) takes Devries et al. (2011) as a starting point but has several differences. The most explicit ones are the extension of the dataset from 2007 to 2014, which is particularly important given the number of austerity policies implemented in this period, and the exclusion of the Netherlands from the sample. However, the changes are deeper, as the authors re-classify the shocks based on the original sources and, thus, significant discrepancies

[21] The analyzes is initially presented in Alesina, Favero, et al. (2019a).

Table 2: Description of narrative fiscal shocks in Alesina, Azzalini, et al. (2018)

	Any Size	**>1% GDP**	**>2% GDP**	**>3% GDP**
Total	232	128	69	33
Expansions	9	0	0	0
Contractions	223	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%

Note: Range of shocks is the difference between the largest and smallest shock (in the case with expansions, the largest expansion is considered the 'smallest shock').

in the size of the shocks are frequent, and it is not rare that episodes found in one sample are not present in the other. Some basic descriptive statistics of the sample are displayed in table 2; more than half of the contractionary shocks are smaller than 1% of GDP and only around 15% is larger than 3% of GDP. As indicated before, we use only the negative shocks in the sample.

Alesina, Azzalini, et al. (2018) implement vector autoregressions to evaluate the effects of such shocks. 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.

As indicated in section 2, the 'treatment model' is used to calculate the probability of each unit (country-year) to have an austerity shock. The variables used in the probit to estimate this probability are:[22] country dummies[23], debt (% GDP), GDP gap (as measured by HP filter), 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, Azzalini, et al. (2018).[24] After the 'preliminary'

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

[23]In the appendix, we also test the results including time dummies.

[24]It can be found here: 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 4 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 (ibid.) 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.

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. We follow Jordà and Taylor (2016) and control for a cyclical component of GDP, country-fixed effects, and two lags of change in GDP.[25]

More specifically, the estimator can be written as:

$$\hat{\Lambda}^h_{AIPW} = \frac{1}{n}\sum_t \Bigg\{ \left[\frac{D_t(y_{t+h}-y_t)}{\hat{p}_t} - \frac{(1-D_t)(y_{t+h}-y_t)}{(1-\hat{p}_t)} \right] - \frac{(D_t-\hat{p}_t)}{\hat{p}_t(1-\hat{p}_t)} \left[(1-\hat{p}_t)m^h_1(X_t,\hat{\theta}^h_1) + \hat{p}_t m^h_0(X_t,\hat{\theta}^h_0) \right] \Bigg\} \quad (1)$$

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 be the size of $(y_{t+h} - y_t)$ for group j in the absence of treatment and β^h_j the estimator of the covariates over $(y_{t+h} - y_t)$. [26]

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 1 presents the main results of our estimations, namely the effects on GDP of contractionary fiscal shocks of different sizes. As can be seen, when all contractionary shocks are considered, a negative effect on GDP is present in most years, but in a statistically significant way only in the fourth and fifth years after the shock. The results are different for larger shocks: for those larger than 1.5% of GDP, the coefficients tend to be larger (in absolute terms) and

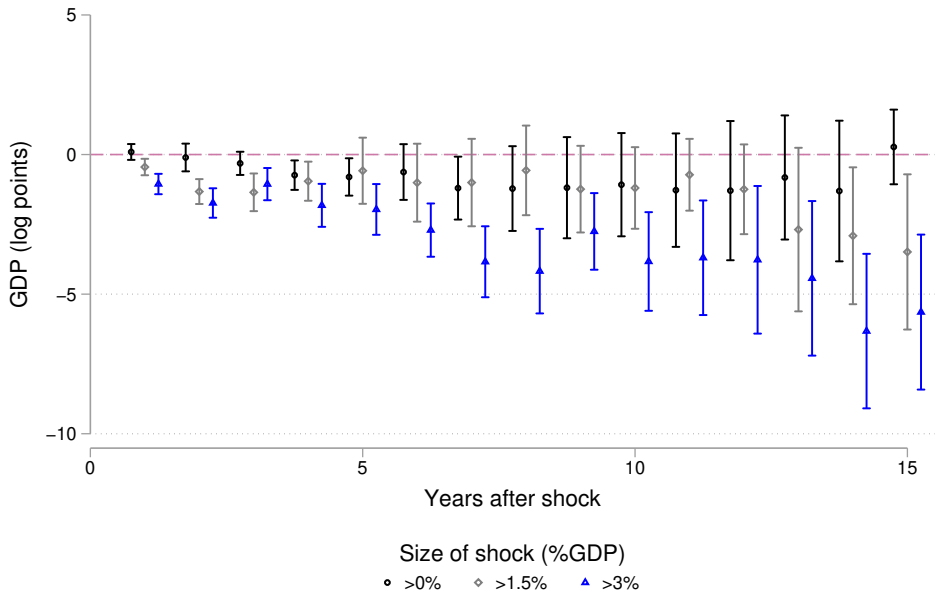
[25] In appendix 6.1, we test with an additional lag of GDP change to address any concern with pre-trends.

[26] 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 $\theta^h_0 = \theta^h_1$.

more significant (statistically), including after 15 years, for which the coefficient is -3.5% of GDP. When restricting the analysis to stronger shocks, larger than 3% of GDP, the coefficients are even more negative and statistically significant for every year; those shocks are associated with a reduction in GDP of 5.6% after 15 years.[27] In other words, our estimations suggest that relatively large contractionary fiscal shocks generate significant long-run negative effects on GDP.

It can be argued that the choice for the thresholds for the minimum size of the shock is somehow subjective. The choice of 1.5% (and its multipliers) of GDP as the baseline follows the threshold adopted in some papers to establish the minimum size of the change in the cyclically adjusted primary balance for a fiscal shock (Alesina and Perotti (1995), Alesina and Ardagna (2010)). To reduce concerns that these choices are driving our results, we test other minimum thresholds in the robustness section (subsection 4.1).

Figure 1: Effect of Austerity - By size of the shock



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

It is important to note that the narrative fiscal data from Alesina, Azzalini, et al. (2018) considers fiscal *plans* and divides the austerity measures into three categories: (i) shocks that took place in time t and were not previously announced; (ii) measures that take place in time t

[27] These results do not tell us nothing about different proportional effects of the shocks; we explore this issue in the next section (subsection 4.4).

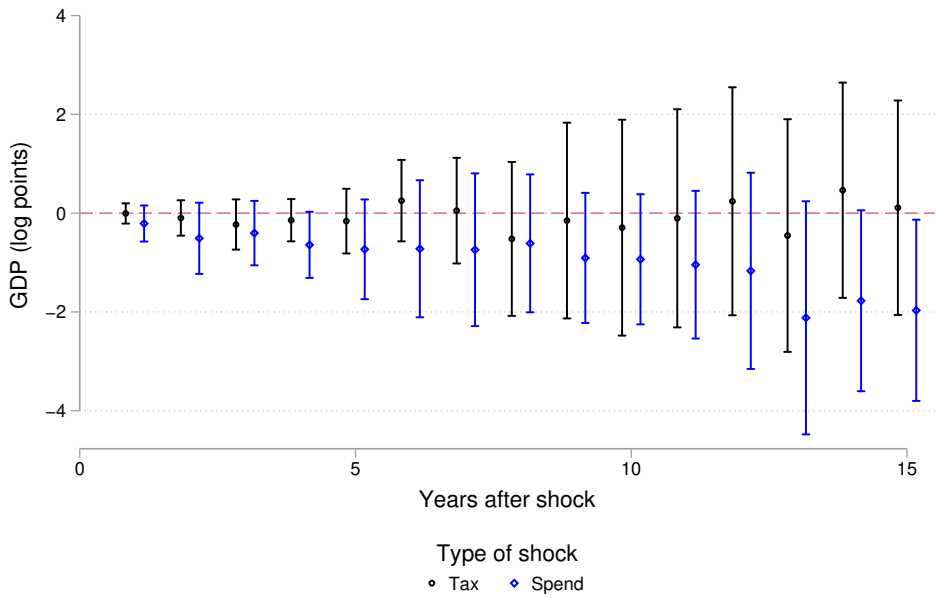
and that were announced in the past; and (iii) measures announced in time t to be implemented in $t+1$. In line with the authors' use of their dataset in Alesina, Azzalini, et al. (2018)[28], the size of a shock in time t is assumed to be the sum of all three categories. In the appendix (section 6.4), we also add robustness exercises of estimations that use only actual shocks (anticipated or not) and the results of negative effects in the long-run persist for shocks larger than 3% of GDP, with even larger negative effects after 15 years.

An important discussion in the literature is whether austerity policies based on expenditure reduction have the same effects as those implemented via tax increases. In figure 2, we explore the question by looking at each type of policy. There are different ways of defining each of these shocks; in the baseline specification used here, we consider all tax or expenditure shocks - a usual alternative in the literature is to consider a tax or expenditure shock if the larger part of the austerity measure is based on it. While tax increases do not generate any significant change in GDP, reductions in government expenditure tend to significantly decrease GDP over extended periods. The result that spending cuts harm GDP more than tax increases is also robust to the use of only actual shocks that take place in time t, as can be seen in the appendix.

Hence, spending-based fiscal shocks seem to be responsible for the significant negative long-run effect of austerity found in our overall estimations. This result differs from the majority obtained in the fiscal multiplier literature that uses time-series data, as summarized by V. Ramey (2019). These papers frequently find a larger negative multiplier (for up to five years) for tax increases than for spending cuts. Effects estimated by calibrated DSGE models are, in general, more in line with the results found here.

[28] See, for instance, the discussion in p. 149 about expenditure vs. tax based shocks. This is also described in the book Alesina, Favero, et al. (2019a).

Figure 2: Type of shock



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

3.2 Extended dataset and different GDP measure

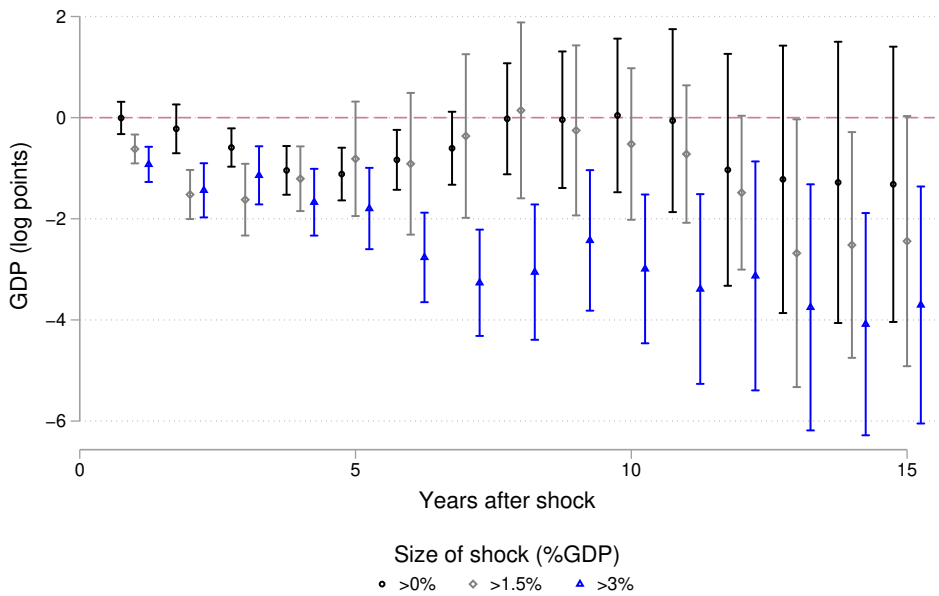
The discussion regarding austerity regained centrality after the great financial crisis of 2007 and its repercussions in the European debt crises some years later. This period was marked by countries adopting fiscal austerity measures with the goals of controlling indebtedness and increasing GDP growth, which could be directly connected due to a "debt intolerance" (Reinhart and Rogoff (2010)) or the channels indicated by the "expansionary austerity" hypothesis presented before.

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 nonexistence of a longer narrative dataset, we perform an intermediate solution: while we keep using the same narrative shocks as treatments, we extend the series of shocks to be used as controls with a measure of fiscal shock based on the cyclically-adjusted primary balance (CAPB) calculated by the IMF. 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 there might have been revisions in the growth rates since the data was compiled by Alesina, Favero, et al. (2019a) - we use data from OECD on the growth rate of GDP (in volume). As a robustness test, available in the appendix, we perform the same estimation using GDP at constant national prices, from PWT 10.0; the results are very similar.

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. [29]

Figure 3: Extension - Alesina, Azzalini, et al. (2018)



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

4 Extensions and Robustness

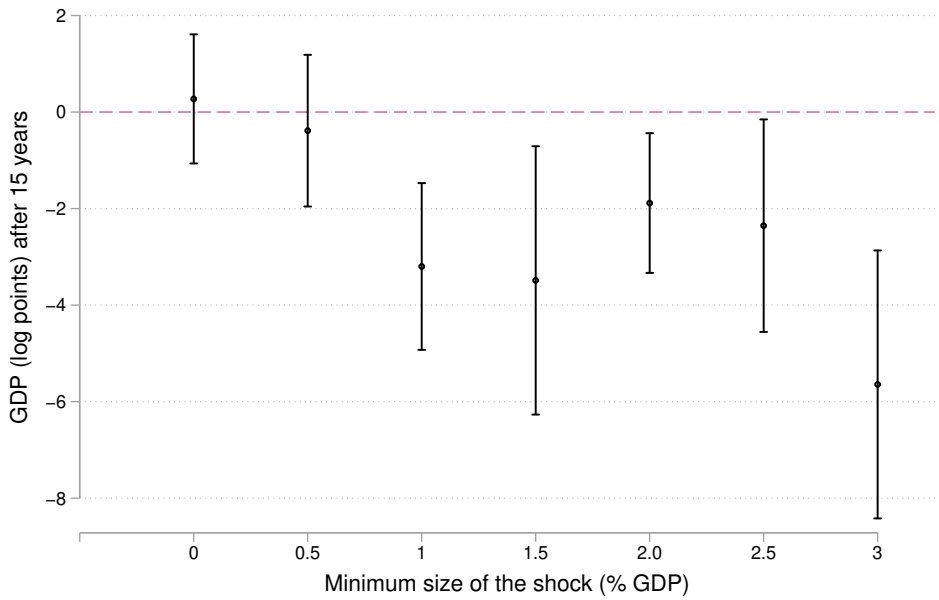
4.1 Alternative thresholds

As indicated above, the choice of 1.5% of GDP as our baseline threshold is based on other important papers in the literature. However, it is clear that if the results are too sensitive to this threshold, the generality of our argument - that austerity shocks, understood as significantly

[29] We do not calculate the effect by type of austerity with this extended dataset as the measure of CAPB that we am using do not discriminate between tax and expenditure changes; therefore, its use, even only as a control and for a short period of time would not be adequate.

large negative fiscal shocks, have long-run effects - is weakened. To address this, we test the effect on GDP after 15 years of shocks considering four other thresholds. Figure 4.1 indicates that shocks larger than 1%, 2%, and 2.5% of GDP - besides the baseline cases of 1.5% and 3% - have long-run effects on GDP.

Figure 4: Alternative thresholds



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

4.2 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. As mentioned before, this dataset has several differences with respect to the one elaborated by Alesina, Favero, et al. (2019a) even for the years covered by both, and it excludes this most recent wave of austerity plans after the Global Financial Crisis. Thus, checking if the effects of this alternative sample of shocks align with our baseline results can serve as an important robustness check.

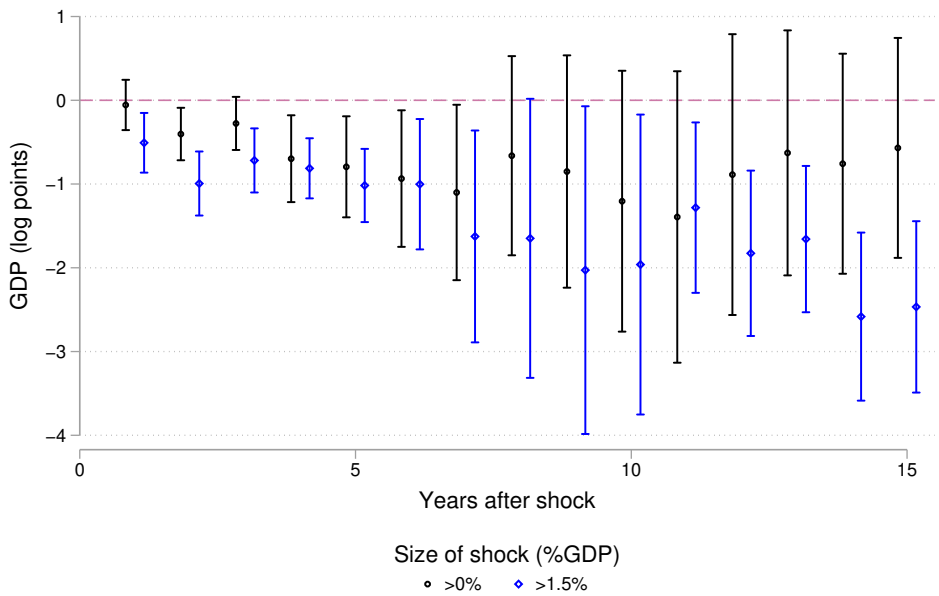
Applying our baseline estimation strategy and taking advantage that this is also the method implemented by Jordà and Taylor (2016), we employ the same data used by them to calculate the probability of being "treated".[30] Only the dummy for an episode of fiscal consolidation in

[30] As indicated in section 3, the variables are: country dummies, debt (% GDP), GDP gap (as measured by HP filter), real GDP growth (current and one lag), long-term and short-term interest rates, current account (% GDP), change in the investment to GDP ratio, real private loan growth, CPI inflation rate, and a dummy for an

the previous year is slightly different from the one of Jordà and Taylor (2016), as they use a treatment variable that also includes fiscal expansions.[31]

Figure 5 presents our results, which are very similar to the ones from Jordà and Taylor (ibid.) 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. Using again our baseline threshold of shocks larger than 1.5% of GDP, the negative effect on GDP is statistically significant in all years after the shock with the exception of the eighth year.

Figure 5: Effect by shock size - alternative dataset



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

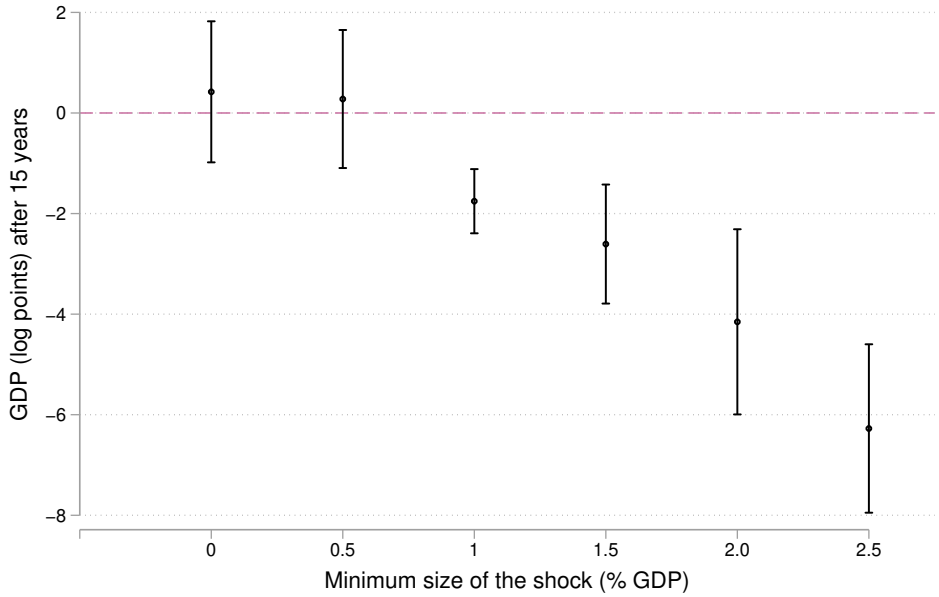
The narrative shocks in the Devries et al. (2011) dataset tend to be smaller than the ones in Alesina, Favero, et al. (2019a), and thus there are not enough observations to perform the estimations for shocks larger than 3% of GDP. In figure 6, we apply the same reasoning used in section 4.1 and get a qualitatively similar result indicating that our findings regarding the

episode of fiscal consolidation in the previous year.

[31] There seems to be a problem in Jordà and Taylor (2016) as it is not the case that the authors are assuming that expansions and contractions are symmetrical, but instead, in their regressions, expansions are entering as contractions - that is, the dummy of treatment takes 1 for both expansions and contractions. That ends up being a minor problem in practical terms given that the number of expansions is very small. However, there are some differences if the estimation is adjusted to contain only contractions: the effect after 5 years, in the restricted case (table 8), drops to -0.9% and is significant only at 10%, while in Jordà and Taylor (ibid.), the effect is of -1.1% and significant at 5%.

long-run effects of austerity shocks are robust to different thresholds for the minimum size of the shocks.

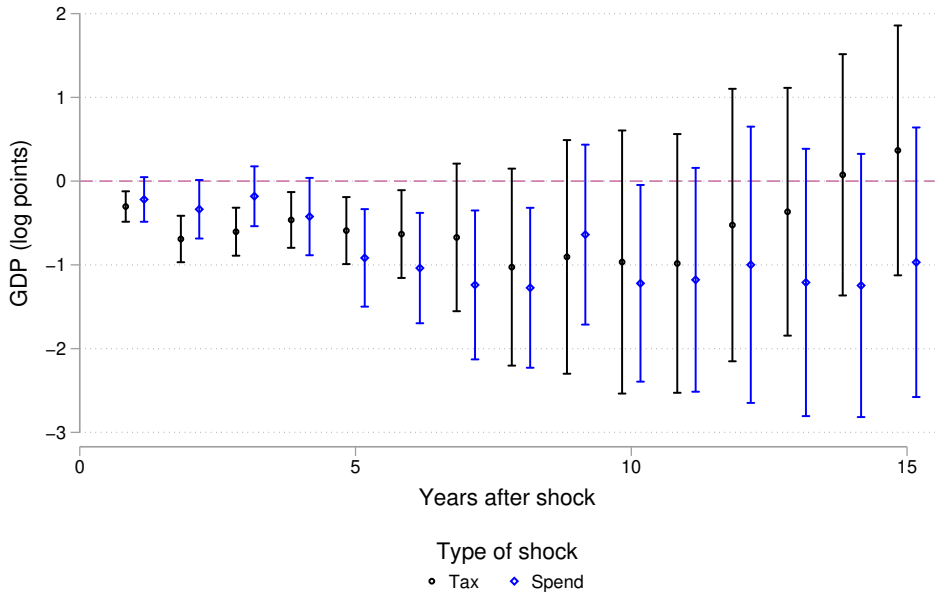
Figure 6: Effect by shock size - alternative dataset and thresholds



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

Still using the data employed by Devries et al. (2011), one can check if the results by the type of shock also hold. Figure 7 indicates that, although expenditure cuts tend to have a more negative effect in the long-run, the coefficients are consistently statistically non-significant at 10% eleven years after the shock.

Figure 7: Effect by type of shock - alternative dataset

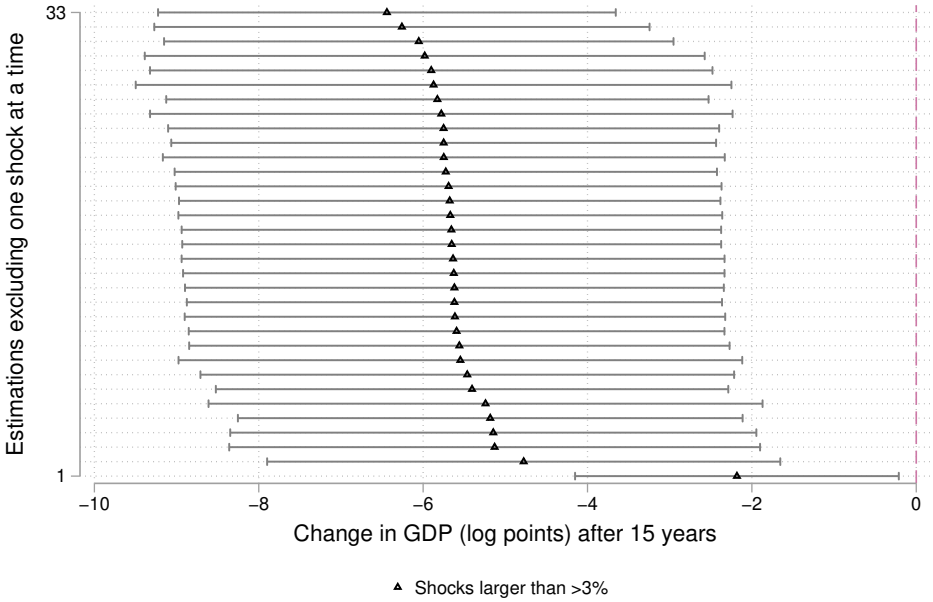


Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

4.3 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 shocks larger than three percent of GDP excluding one episode at a time and check if the effects on GDP after fifteen years hold. Figure 8 shows that the results are robust to the exclusion of any particular shocks.

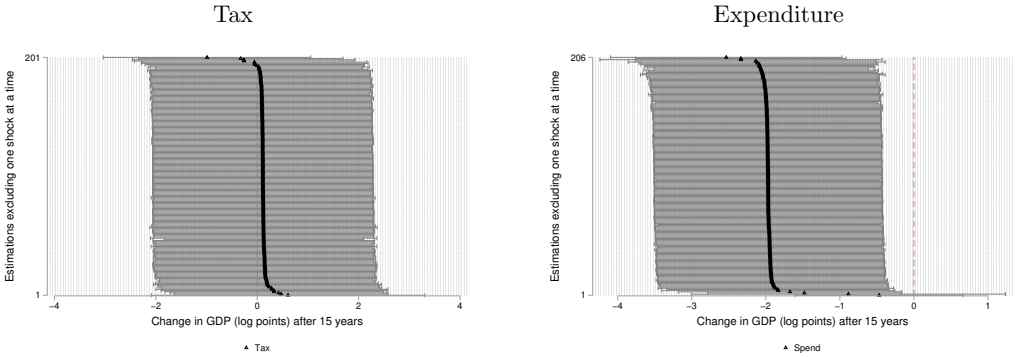
Figure 8: Robustness check - Excluding shocks (Larger than 3% of GDP)



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

The same exercise is performed for the types of shocks. As can be seen in figure 9, the result that increases in taxes are not associated with a change in GDP is consistent with the exclusion of any particular shock. For the case of spending, all estimations indicate a large negative coefficient, although the exclusion of three particular shocks decreases the statistical significance - in two of them, notwithstanding, it remains significant at 90%.

Figure 9: Robustness check - Excluding shocks - Type

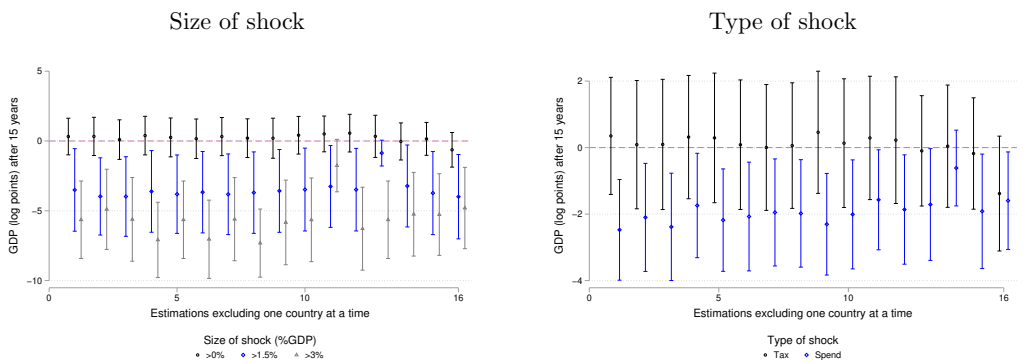


Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

A final exercise excludes entire countries of 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, Favero, 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 10, the effects by size and type of shock are very similar to the baseline estimation, with austerity measures larger than 1.5% and 3% of GDP having a negative and statistically significant effect after 15 years in the vast majority of cases, as well as the effect of expenditure cuts in contrast to increases in taxes.

Figure 10: Robustness check - Excluding countries



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

4.4 Continuous treatment

A question that might be of interest is if the proportional effects of the shocks of different sizes are also relevant. 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 re-weight 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.[32] This is performed within each treatment band of interest of our baseline estimation: all contractionary shocks, and those larger than 1.5% and 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% of GDP will reduce GDP in around 3%

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

Table 3: 'Multipliers' - by size shock

	>0% GDP	**>1.5% GDP**	**>3% GDP**
Instantaneous (after 1 year)			
Multiplier	-0.07	-0.24	-0.23
P-value = >0% GDP	-	0.00	0.02
Long-run (after 15 years)			
Multiplier	-0.51	-1.46	-1.45+
P-value = >0% GDP	-	0.00	0.03

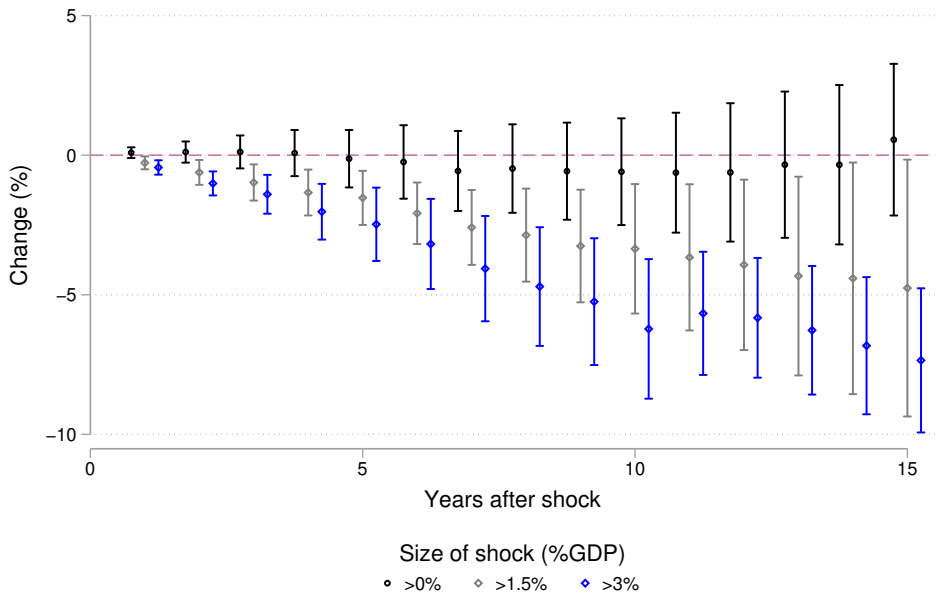
Note:A qui-square test is used to test the null hypothesis that the multiplier is equal to the one when all contractionary shocks are considered. + indicates statistical significance at 10%).

after 15 years. One interesting result is that the multipliers for shocks larger than 1.5% and 3% of GDP are very similar in both short- and long-runs. However, the most important result is that the multipliers for these sufficiently large shocks are significantly different than the one when considering all fiscal contractions; using a qui-square test, we can reject the hypothesis that they are statistically equal with a 5% significance level. This 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.

4.5 Initial examination of channels

A detailed examination of the channels through which these long-run effects operate is beyond the scope of this paper. Taking advantage of readily available data, however, we perform a first approximation to check the effects on the two main aggregate inputs: capital stock and labor. Figure 11 suggests that austerity shocks larger than 1% of GDP (using our baseline dataset) are associated with a consistent and statistically significant negative effect on the stock of capital (as measured by the PWT 10.0).

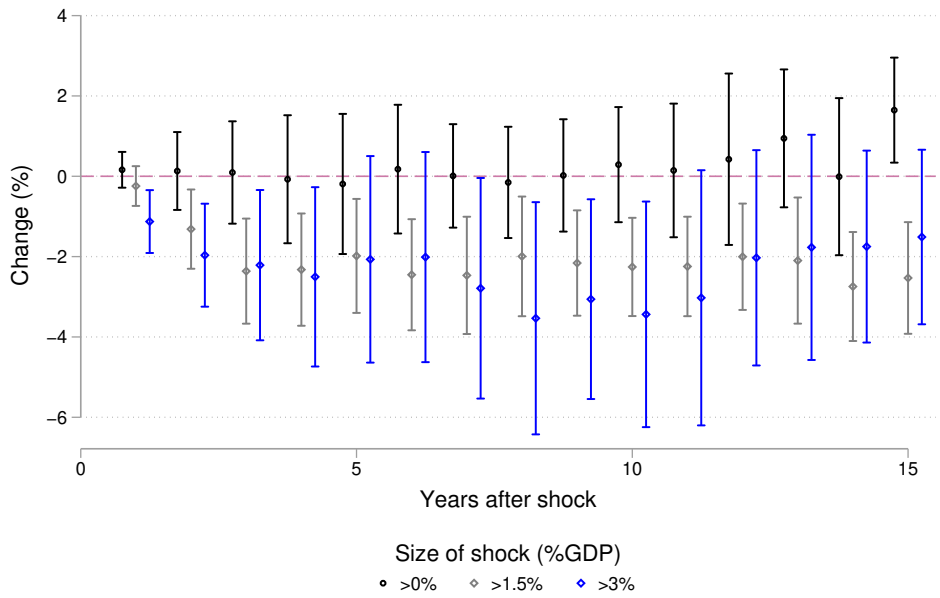
Figure 11: Effect by size - Stock of Capital



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

The effects on the labor market are less clear: in figure 12, larger shocks seem associated with a decrease in the ratio of the employed populated (also calculated based on PWT 10.0 data). This measure, although with advantages in some dimensions (do not rely on different definitions of unemployment and incorporate changes in the population actively searching for work, for instance), also has clear drawbacks, such as a change in demographics over fifteen years. Moreover, results for other measures, such as the short and long-run unemployment rates and the labor force participation as measured by the OECD, as presented in the appendix, do not present a clear picture.

Figure 12: Effect by size - Employed population



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

4.6 Simpler Difference-in-Differences method

Another relevant exercise is to check if the results are too sensitive to our method. On the one hand, the baseline method is chosen as it is the most appropriate to estimate the effects of interest here, given the arguments presented in section 2. In this sense, it is *expected* that the estimated effects depend on the method. On the other hand, if the results are reverted with the use of other methods, although the baseline results should not be discarded, one would need to analyze in greater detail the assumptions made in our baseline method and why the results differ.

In this subsection, thus, we perform the analysis with a simpler estimation: instead of weighting the sample using IPW, we control for all the variables in a standard difference-in-differences setup:

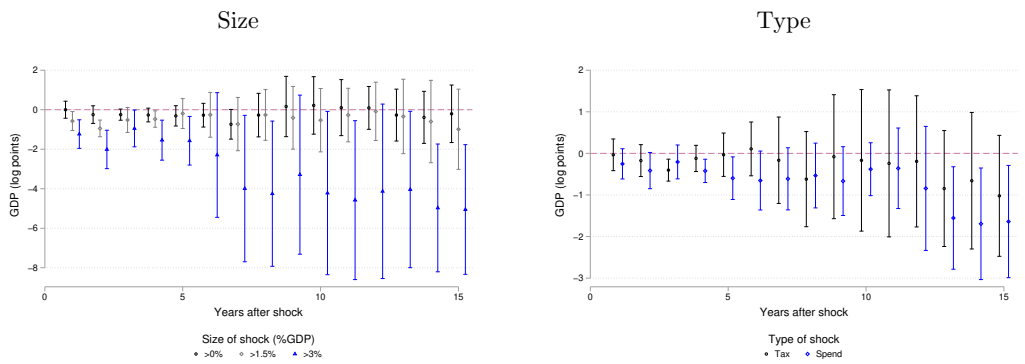
$$\Delta y_{i,t+h} = \alpha^i + \beta^s E_{i,t} + \theta X_{i,t} + \epsilon_{i,t+h} \tag{2}$$

where α^i are country dummies and X, a vector with all the control variables, including those used in both the "treatment" and "outcome" models in the AIPW estimator[33].

[33]The controls are: debt (% GDP), GDP gap (as measured by HP filter), 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, CPI inflation rate,

As can be seen in figure 13, the results are, in general, in line with the estimations using the AIPW method, the main differences being the smaller coefficient for shocks larger than 1.5% of GDP and the larger confidence interval for estimations of the shocks larger than 3% of GDP. The effects of the types of shocks are very similar to our baseline estimations.

Figure 13: Simpler DiD



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval. Actual shocks are those that took place in time t, expected or not.

4.7 "Cleaner" controls - A local projections approach to DiD

An increasingly recognized problem in studies that resort to some form of differences-in-differences estimation is the bias that emerges once one moves away from a "2X2" setup - that is, two periods (pre and post-treatment) and two status (treated or never treated) (e.g.,Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020) and 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 subsection, 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 $t+k$.[34] This is performed with our baseline setting (section

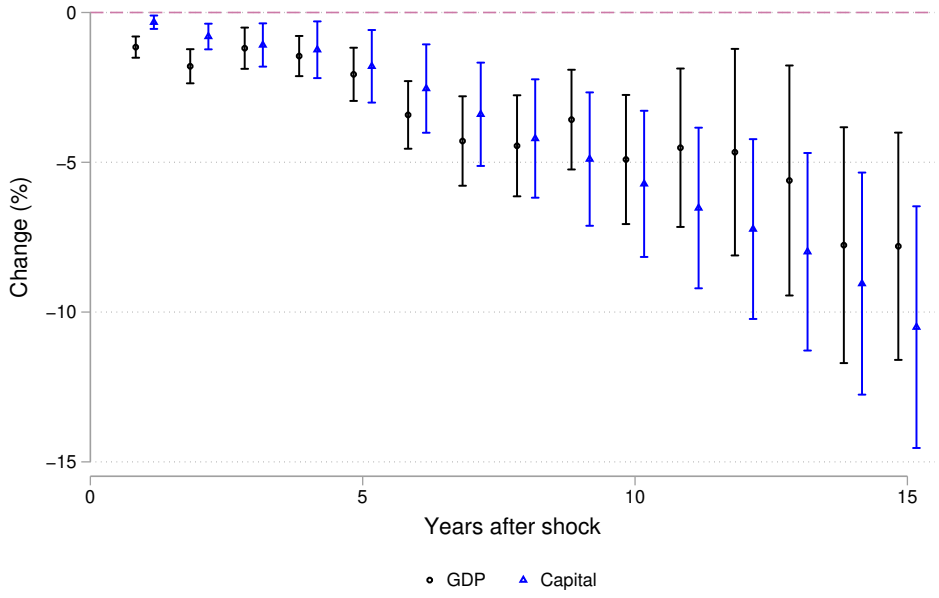
a cyclical component of GDP, country fixed effects, two lags of change in GDP, and leads of shocks that occur from time t up to time h.

[34]For instance, assume several countries have austerity shocks in 1990. To calculate the average effect of these

3), 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.[35] We focus, therefore, on the higher threshold of shocks larger than 3% of GDP so that we can have an adequate number of observations. Results for both GDP and capital stock are displayed in figure 14. As can be seen, even in this much stricter scenario, results persist, indicating significant negative effects of austerity shocks over long periods.

Figure 14: Cleaner controls - Shocks larger than 3% of GDP



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

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 5 years, the control sample would consist of countries that did not experience an episode between 1990 and 1995.

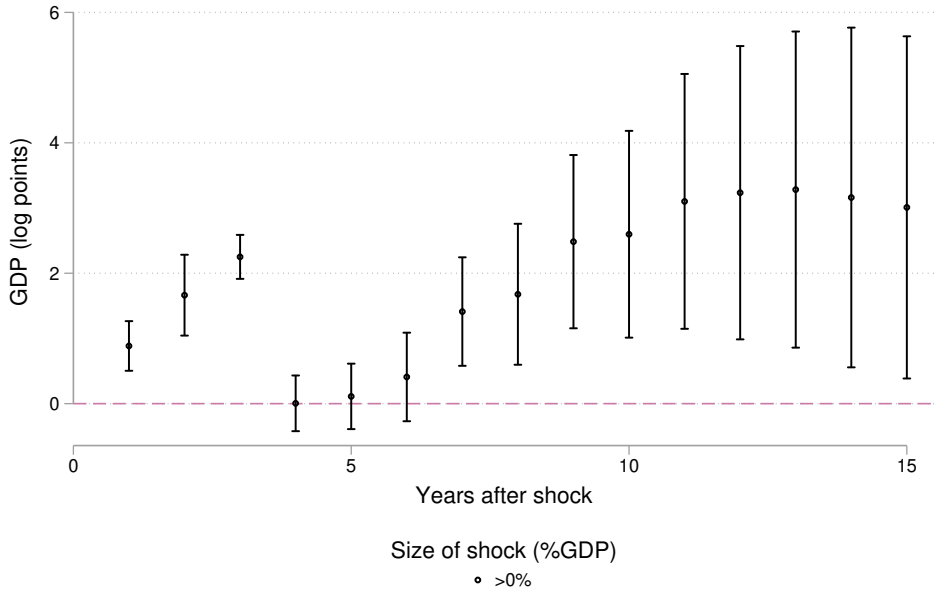
[35] That is, for a smaller threshold, we have a larger number of shocks, and thus the number of countries that can be used as controls in a 15-years window is very reduced.

4.8 Expansions

The goal of this paper is to analyze the long-run effects of austerity shocks. However, differently from the existing literature, by focusing on the *actual* negative fiscal shocks and not assuming that positive shocks are symmetrical, we can also perform an initial assessment of the effects of expansionary fiscal measures. These shocks are much rarer in the existing narrative datasets, amounting to only nine cases in our baseline one (Alesina, Favero, et al. (2019a)), and, therefore, these results must be interpreted with all the due caveats and should be seen only as a first approximation to the issue.

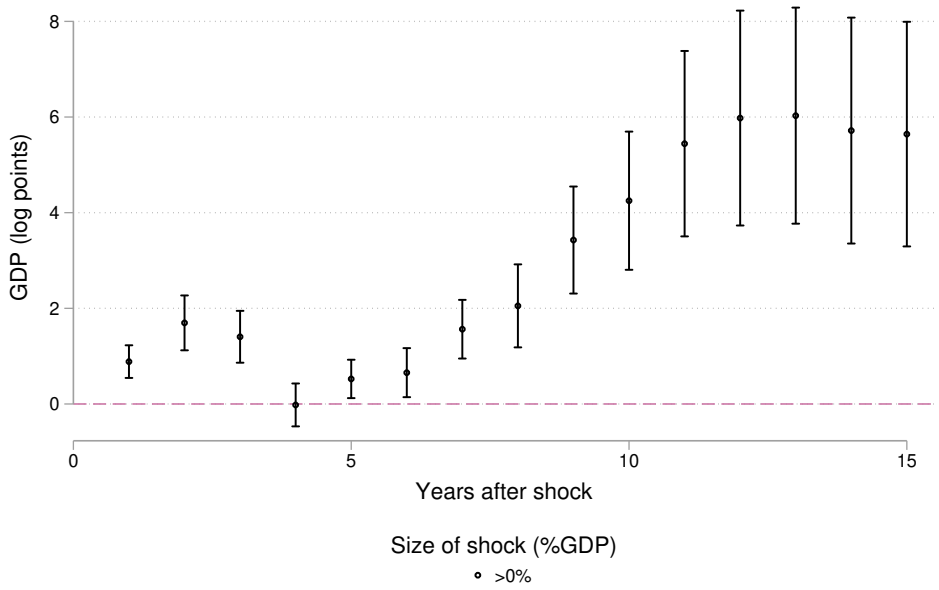
Using the same regression as in section 3 and both our baseline dataset (figure 15) and its extended version (figure 16), we find that expansionary shocks tend to have positive long run effects on GDP.

Figure 15: Effect of fiscal expansions - baseline



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

Figure 16: Effect of fiscal expansions - extended dataset



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

5 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 at addressing one in particular: the long-run effects of austerity policies.

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 focused, in general, to a maximum of five years, even on its more aggregated level, such as output or capital stock.

Employing a method that 're-randomize' the allocation of austerity episodes in a local pro-

jections setup and accounting for the fact that multiple shocks occur in the time horizon of interest, our results indicate that relatively large austerity measures have detrimental effects on GDP 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 countries, to the implementation of simpler regression methods, to the use of 'cleaner' controls, and to a different definition of shocks (only actual shocks). Moreover, there is robust evidence that austerity shocks have significant negative effects on capital stock. There is also some indication, although less robust, of negative effects on the labor market and that spending cuts are more detrimental to GDP than tax increases.

This paper fills a relevant gap in the literature by: (i) examining the long-run effects of fiscal policy, employing techniques that are appropriate for such estimations; (ii) focusing exclusively on contractions and not assuming symmetry with expansions; (iii) allowing different effects for different shock sizes, both in proportional terms and related to its persistence over time. These two last points are particularly relevant as the term 'austerity' is of public interest and it seems 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, do 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 analyze the the long-run impact of fiscal shocks. In this context, our estimations present additional evidence that demand shocks may have significant long-run effects.

References

Acemoglu, D. et al. (2019). "Democracy does cause growth". In: *Journal of Political Economy* 127.1, pp. 47–100.

Agnello, L. and Sousa, R. M. (2014). "How does fiscal consolidation impact on income inequality?" In: *Review of Income and Wealth* 60.4, pp. 702–726.

Alesina, A. and Ardagna, S. (2010). "Large changes in fiscal policy: taxes versus spending". In: *Tax policy and the economy* 24.1, pp. 35–68.

Alesina, A., Azzalini, G., et al. (2018). "Is it the "How" or the "When" that Matters in Fiscal Adjustments?" In: *IMF Economic Review* 66.1, pp. 144–188.

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

— (2019b). "Effects of austerity: Expenditure-and tax-based approaches". In: *Journal of Economic Perspectives* 33.2, pp. 141–62.

Alesina, A. and Perotti, R. (1995). "Fiscal expansions and adjustments in OECD countries". In: *Economic policy* 10.21, pp. 205–248.

Auerbach, A. J. and Gorodnichenko, Y. (2012a). "Fiscal Multipliers in Recession and Expansion". In: *Fiscal Policy after the Financial Crisis.* University of Chicago Press, pp. 63–98.

— (2012b). "Measuring the output responses to fiscal policy". In: *American Economic Journal: Economic Policy* 4.2, pp. 1–27.

Babcock, L. and Loewenstein, G. (1997). "Explaining bargaining impasse: The role of self-serving biases". In: *Journal of Economic Perspectives* 11.1, pp. 109–126.

Baccaro, L., Bremer, B., and Neimanns, E. (2021). "Till austerity do us part? A survey experiment on support for the euro in Italy". In: *European Union Politics* 22.3, pp. 401–423.

Ball, L. M. et al. (2013). *The distributional effects of fiscal consolidation.* International Monetary Fund.

Baum, M. A., Poplawski-Ribeiro, M. M., and Weber, A. (2012). *Fiscal Multipliers and the State of the Economy.* International Monetary Fund.

Blanchard, O., Cerutti, E., and Summers, L. (2015). *Inflation and activity–two explorations and their monetary policy implications.* Tech. rep. National Bureau of Economic Research.

Blanchard, O. and Perotti, R. (2002). "An empirical characterization of the dynamic effects of changes in government spending and taxes on output". In: *the Quarterly Journal of economics* 117.4, pp. 1329–1368.

Blanchard, O. and Quah, D. (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances". In: *American Economic Review* 79.4, pp. 655–673.

Callaway, B. and Sant'Anna, P. H. (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

Cerra, V. and Saxena, S. C. (2008). "Growth dynamics: the myth of economic recovery". In: *American Economic Review* 98.1, pp. 439–57.

Cover, J. P. (1992). "Asymmetric effects of positive and negative money-supply shocks". In: *The Quarterly Journal of Economics* 107.4, pp. 1261–1282.

David, A. and Leigh, D. (2018). *A New Action-based Dataset of Fiscal Consolidation in Latin America and the Caribbean*. International Monetary Fund.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". In: *American Economic Review* 110.9, pp. 2964–96.

De Long, J. B. et al. (1988). "How does macroeconomic policy affect output?" In: *Brookings Papers on Economic Activity* 1988.2, pp. 433–494.

Devries, P. et al. (2011). "A new action-based dataset of fiscal consolidation". In.

Dube, A. et al. (2022). "A local projections approach to difference-in-differences event studies". In: *NBER Working Paper*.

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

Fehr, E. and Tyran, J.-R. (2001). "Does money illusion matter?" In: *American Economic Review* 91.5, pp. 1239–1262.

Fetzer, T. (2019). "Did austerity cause Brexit?" In: *American Economic Review* 109.11, pp. 3849–86.

Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.

Greenwald, B. C. et al. (1988). "Examining alternative macroeconomic theories". In: *Brookings Papers on Economic Activity* 1988.1, pp. 207–270.

Guajardo, J., Leigh, D., and Pescatori, A. (2014). "Expansionary austerity? International evidence". In: *Journal of the European Economic Association* 12.4, pp. 949–968.

Guriev, S. and Papaioannou, E. (2022). "The political economy of populism". In: *Journal of Economic Literature* 60.3, pp. 753–832.

Haltmaier, J. (2013). "Do recessions affect potential output?" In: *FRB International Finance Discussion Paper* 1066.

Ilzetzki, E., Mendoza, E. G., and Végh, C. A. (2013). "How big (small?) are fiscal multipliers?" In: *Journal of monetary economics* 60.2, pp. 239–254.

Jordà, Ò. (2005). "Estimation and inference of impulse responses by local projections". In: *American Economic Review* 95.1, pp. 161–182.

Jordà, Ò., Singh, S. R., and Taylor, A. M. (2020). *The long-run effects of monetary policy.* Tech. rep. National Bureau of Economic Research.

Jordà, Ò. and Taylor, A. M. (2016). "The time for austerity: estimating the average treatment effect of fiscal policy". In: *The Economic Journal* 126.590, pp. 219–255.

Li, D., Plagborg-Møller, M., and Wolf, C. K. (2022). *Local projections vs. vars: Lessons from thousands of dgps.* Tech. rep. National Bureau of Economic Research.

Lucas Jr, R. E. (2003). "Macroeconomic priorities". In: *American economic review* 93.1, 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". In: *Statistics in medicine* 23.19, pp. 2937–2960.

Martin, R., Munyan, T., and Wilson, B. A. (2015). "Potential output and recessions: are we fooling ourselves?" In: *FRB International Finance Discussion Paper* 1145.

Mountford, A. and Uhlig, H. (2009). "What are the effects of fiscal policy shocks?" In: *Journal of applied econometrics* 24.6, pp. 960–992.

Ponticelli, J. and Voth, H.-J. (2020). "Austerity and anarchy: Budget cuts and social unrest in Europe, 1919–2008". In: *Journal of Comparative Economics* 48.1, pp. 1–19.

Rajmil, L. et al. (2020). "Austerity policy and child health in European countries: a systematic literature review". In: *BMC public health* 20.1, pp. 1–9.

Ramey, V. (2019). "Ten years after the financial crisis: What have we learned from the renaissance in fiscal research?" In: *Journal of Economic Perspectives* 33.2, 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, pp. 145–194.

Ramey, V. A. (2011). "Identifying government spending shocks: It's all in the timing". In: *The Quarterly Journal of Economics* 126.1, pp. 1–50.

Ramey, V. A. and Zubairy, S. (2018). "Government spending multipliers in good times and in bad: evidence from US historical data". In: *Journal of Political Economy* 126.2, pp. 850–901.

Reinhart, C. M. and Rogoff, K. S. (2010). "Growth in a Time of Debt". In: *American economic review* 100.2, pp. 573–78.

Riera-Crichton, D., Vegh, C. A., and Vuletin, G. (2016). "Tax multipliers: Pitfalls in measurement and identification". In: *Journal of Monetary Economics* 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". In: *American Economic Review* 100.3, pp. 763–801.

Solow, R. M. (1956). "A contribution to the theory of economic growth". In: *The quarterly journal of economics* 70.1, pp. 65–94.

Summers, L. H. (1988). "Relative Wages, Efficiency Wages, and Keynesian Unemployment". In: *The American Economic Review* 78.2, pp. 383–388.

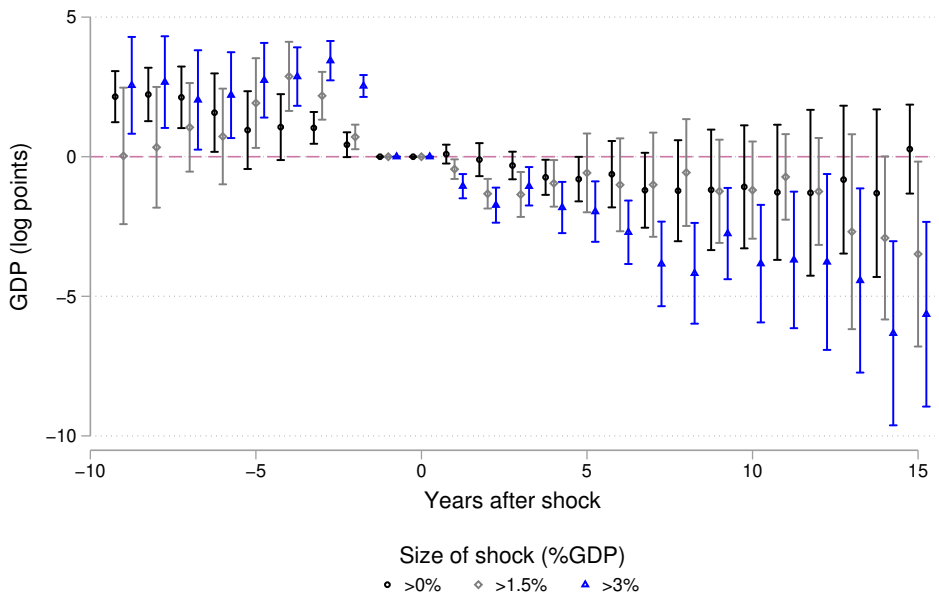
Teulings, C. and Zubanov, N. (2014). "Is economic recovery a myth? Robust estimation of impulse responses". In: *Journal of Applied Econometrics* 29.3, pp. 497–514.

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*. Tech. rep. Board of Governors of the Federal Reserve System (US).

6 Appendix

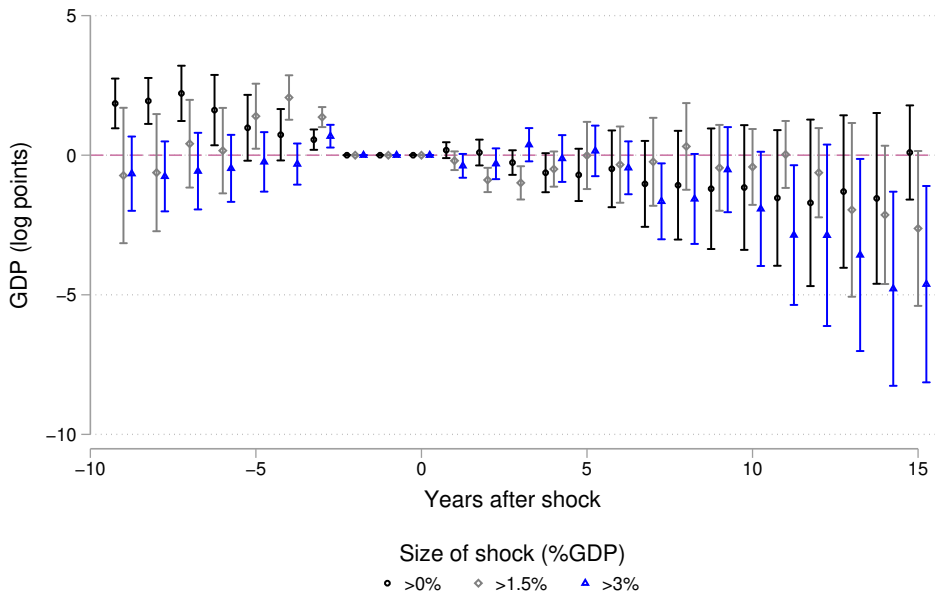
6.1 Pre-trends

Figure 17: Alesina - pre-trends



In our baseline estimation, we follow the setting proposed by Jordà and Taylor (2016). Figure 17 indicates that pre-trends might be operating, which might indicate that, even reweighting the sample, some differences remain between treated and untreated units. As a robustness exercise, we run the same estimation but control for one additional lag of GDP growth (in the 'outcome' model). Figure 18 indicates that this reduces significantly pre-trends within a 10-years window (particularly for the shocks that we are most interested in) while the main results persist.

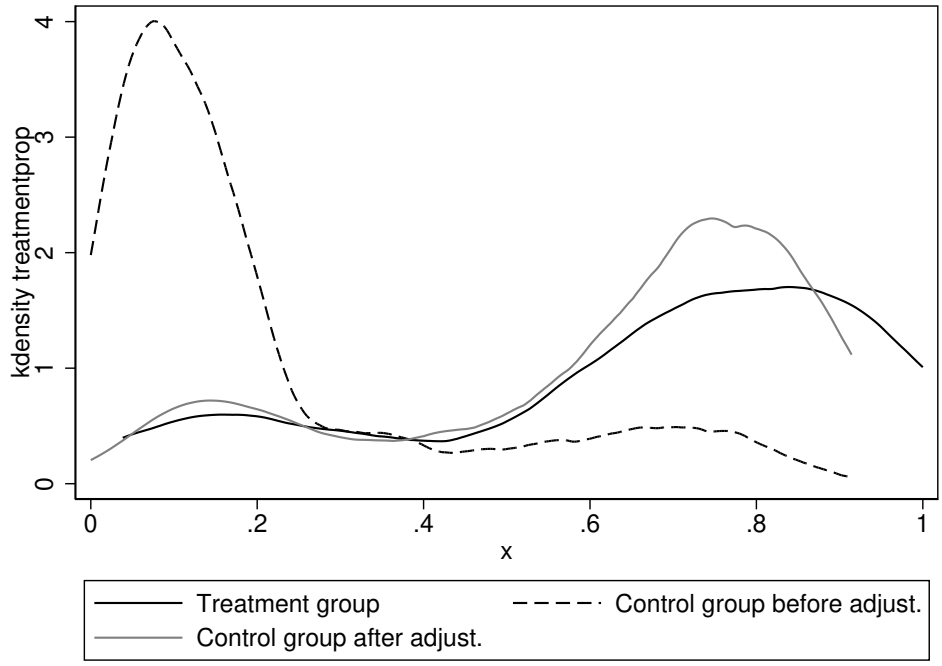
Figure 18: Alesina - pre-trends



6.2 Overlapping assumption

Some details on the data for the Alesina: Debt is the General Government Debt/GDP ratio - IMF Historical Public Debt Database. CPI inflation rate (IMF) Short-term interest rate - OECD Economic Outlook. Long-term interest rate: long-term interest rate on government bonds - OECD Economic Outlook. Investment to GDP ratio based on Gross Fixed Formation data (OECD Economic Outlook) and Gross domestic product, volume, market prices. Source for all except IRL: OECD Economic Outlook n. 97. Source for IRL: IMF WEO April 2015.

Figure 19: Alesina (all austerity cases)



6.3 Extension

Using the OECD data for the capb (2) and PWT for GDP (1)

Figure 20: Extension using IMF CAPB and PWT data for GDP - Alesina, Azzalini, et al. (2018)

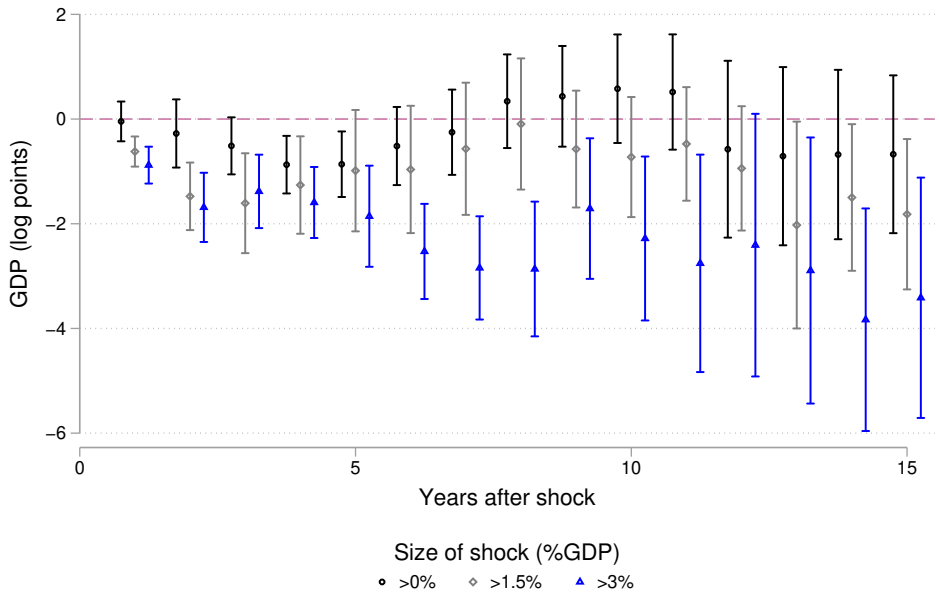
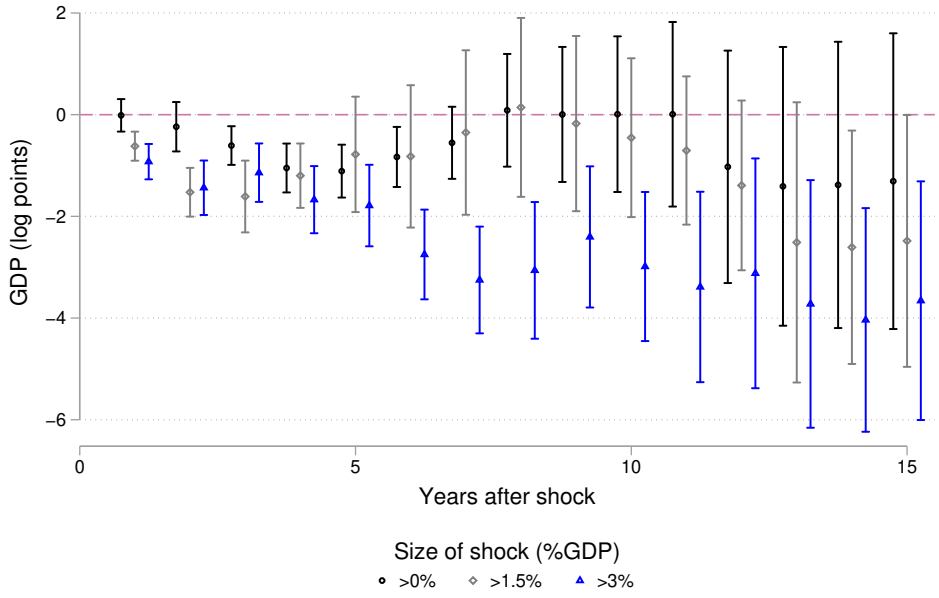


Figure 21: Extension OECD CAPB - Alesina, Azzalini, et al. (2018)



6.4 Only actual shocks

Figure 22: Baseline - Only actual shocks

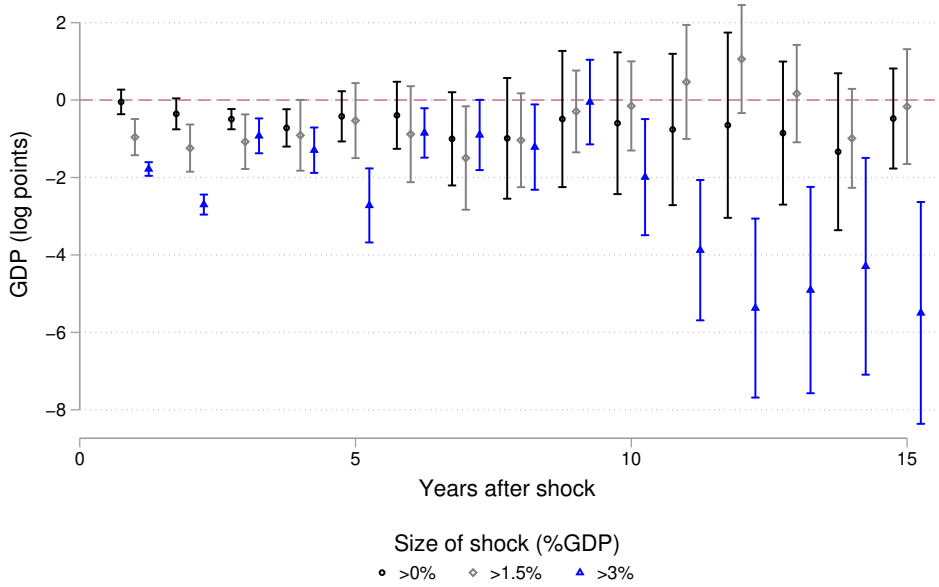


Figure 23: By type of shock - Only actual shocks

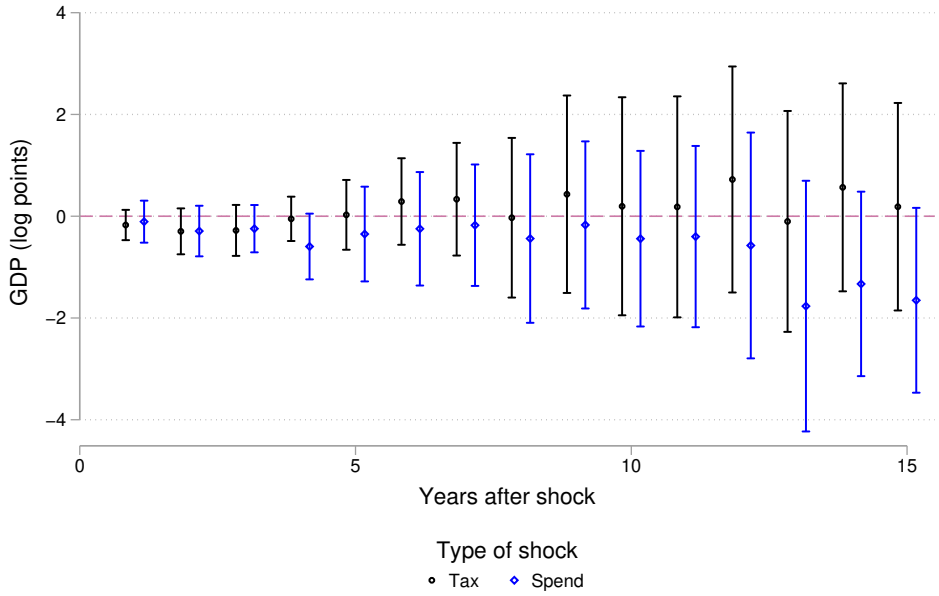


Figure 24: Only actual shocks - Robustness - Size

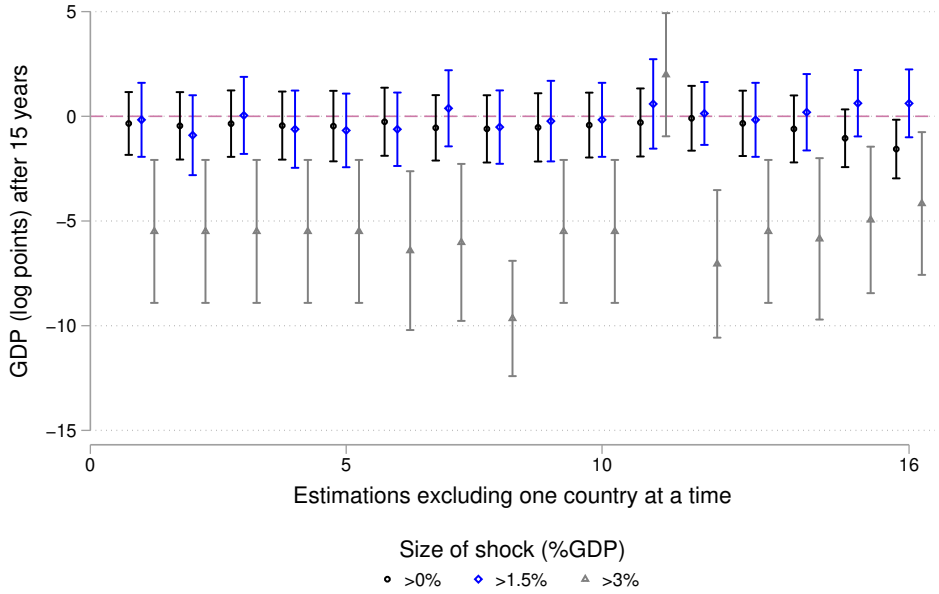


Figure 25: Only actual shocks - Robustness - Type

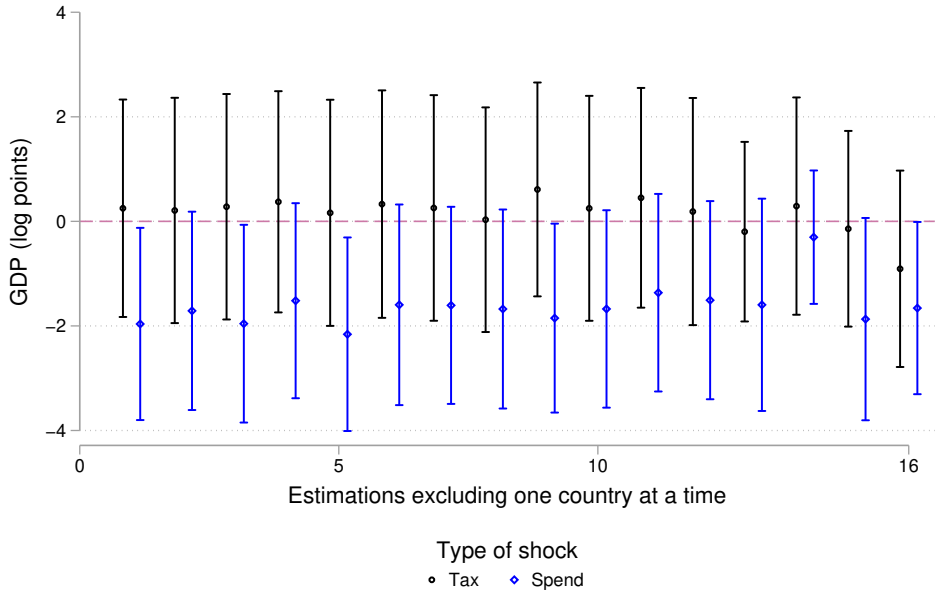
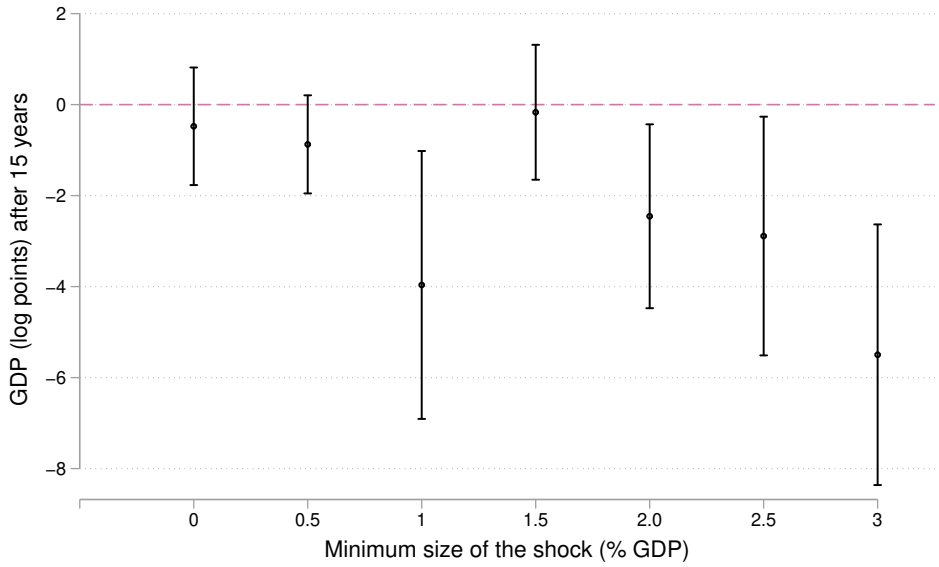
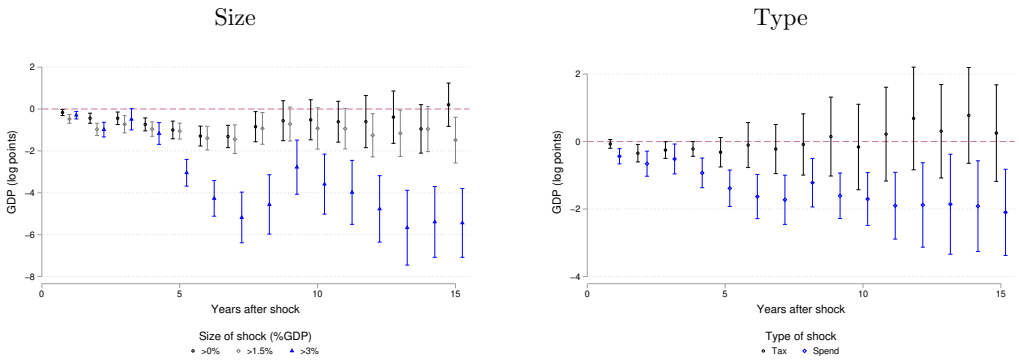


Figure 26: Only actual shocks - By size of shock



6.5 Different fixed effects assumptions

Figure 27: Country and Time fixed effects



Note: Dots indicate estimated coefficients. The bars indicate a 95% confidence interval.

6.6 Effects on the labor market

Figure 28: Effect by size - Unemployment rate - OECD data

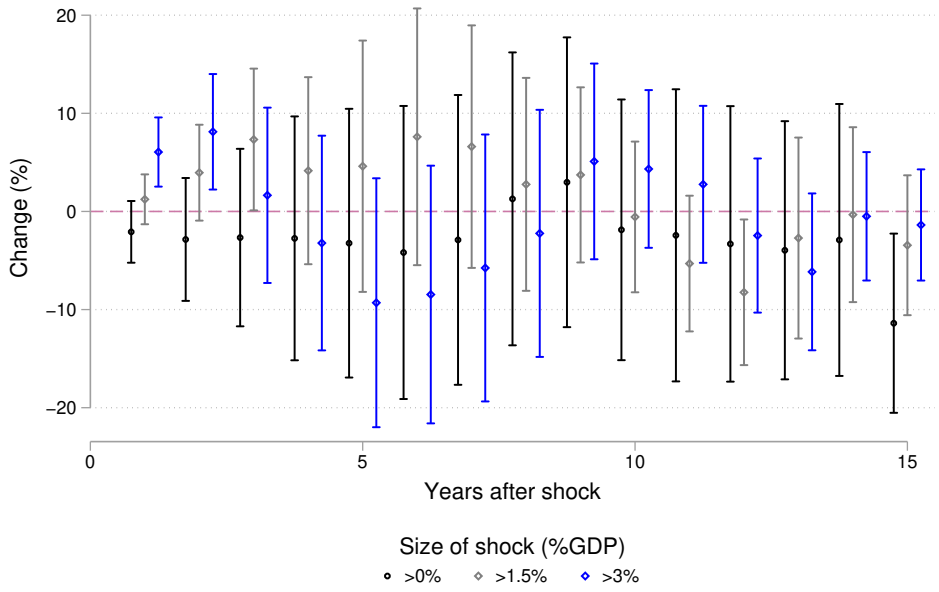


Figure 29: Effect by size - Long-run Unemployment rate - OECD data

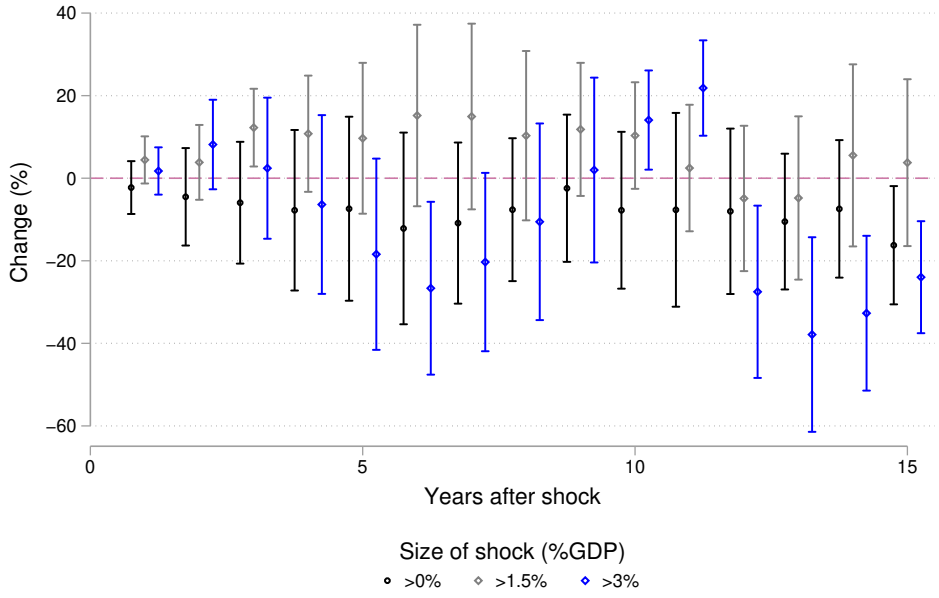
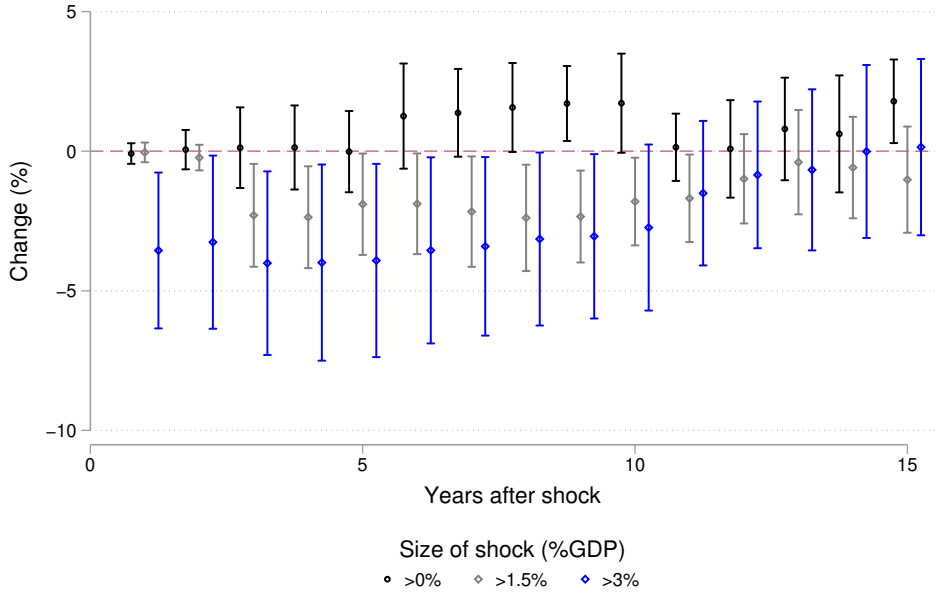
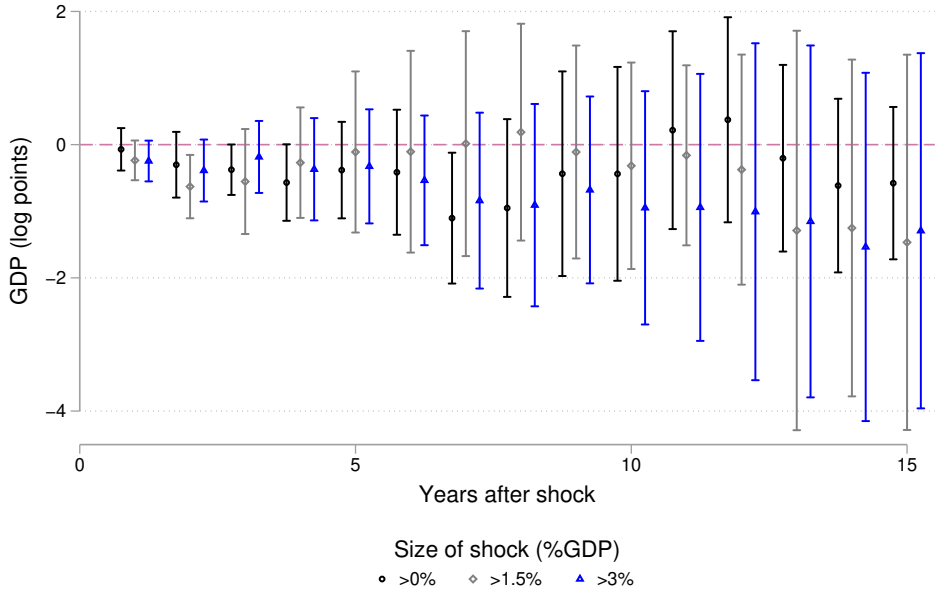


Figure 30: Effect by size - Labor force participation - OECD data



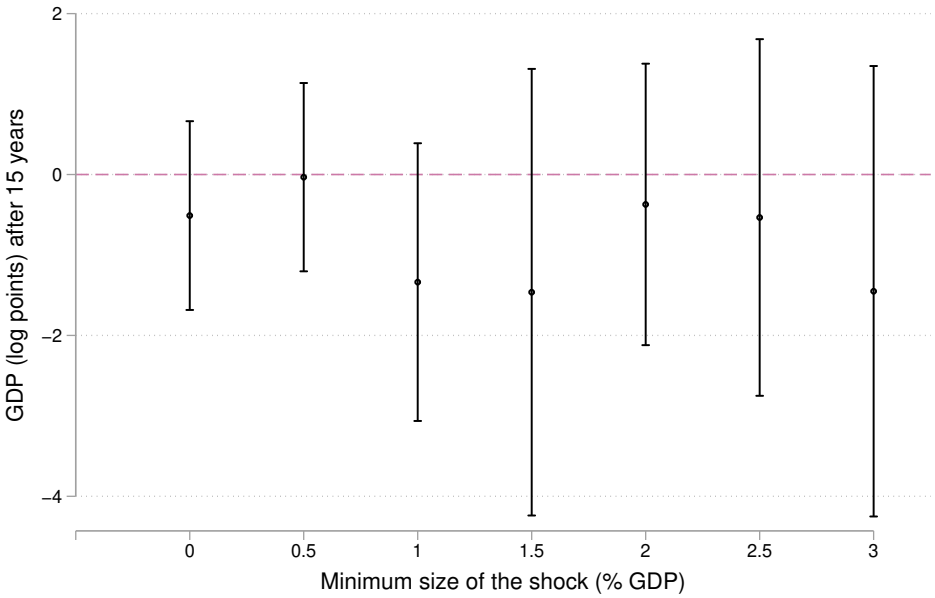
6.7 Multiplier

Figure 31: Effect of Austerity - Multiplier - Extended Dataset (CAPB)



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.

Figure 32: Effect of Austerity - By size of the shock - Multiplier



Note: Dots indicate estimated coefficients. The bars indicate a 90% confidence interval.