WORKING PAPER

19-12 Aggregate Effects of Budget Stimulus: Evidence from the Large Fiscal Expansions Database

Jérémie Cohen-Setton, Egor Gornostay, and Colombe Ladreit July 2019

Abstract

This paper estimates the effects of fiscal stimulus on economic activity using a novel database on large fiscal expansions for 17 OECD countries for the period 1960–2006. The database is constructed by combining the statistical approach to identifying large shifts in fiscal policy with narrative evidence from contemporaneous policy documents. When correctly identified, large fiscal stimulus packages are found to have strong and persistent expansionary effects on economic activity, with a multiplier of 1 or above. The effects of stimulus are largest in slumps and smallest in booms.

IEL Codes: E6, H3, H5, H6, N1

Keywords: Fiscal Policy, Public Economics, Public Finance, Tax Elasticities, National Government Expenditure, National Budget, Macroeconomic Policy, Stabilization, Macroeconomic History

Jérémie Cohen-Setton is research fellow at the Peterson Institute for International Economics. **Egor Gornostay** is research statistician and quality control coordinator at the Peterson Institute for International Economics. **Colombe Ladreit**, former research analyst at the Peterson Institute for International Economics, is a PhD student at Bocconi University.

© Peterson Institute for International Economics. All rights reserved.

This publication has been subjected to a prepublication peer review intended to ensure analytical quality. The views expressed are those of the authors. This publication is part of the overall program of the Peterson Institute for International Economics, as endorsed by its Board of Directors, but it does not necessarily reflect the views of individual members of the Board or of the Institute's staff or management.

The Peterson Institute for International Economics is a private nonpartisan, nonprofit institution for rigorous, intellectually open, and indepth study and discussion of international economic policy. Its purpose is to identify and analyze important issues to make globalization beneficial and sustainable for the people of the United States and the world, and then to develop and communicate practical new approaches for dealing with them. Its work is funded by a highly diverse group of philanthropic foundations, private corporations, and interested individuals, as well as income on its capital fund. About 35 percent of the Institute's resources in its latest fiscal year were provided by contributors from outside the United States. A list of all financial supporters is posted at https://piie.com/sites/default/files/supporters.pdf.

1 Introduction

When the recent financial crisis hit and interest rates fell to zero, policymakers turned to fiscal policy to stimulate a weak economy. Trying to predict the effects of the stimulus, they were surprised to learn that there was a lack of consensus not only about the size of the effects of fiscal policy, but sometimes even about its sign (IMF 2013). Notwithstanding a renaissance of fiscal research since then (Ramey 2019), the range of estimates for what economists call the multiplier remains 'awfully wide' (Blanchard, Leandro, Merler, and Zettelmeyer 2018) for fiscal expansions, where progress in understanding has been more limited than for fiscal consolidations.

A recent strand of the literature has advanced the idea that the multiplier may in fact be smaller for stimulus than for austerity (Barnichon and Matthes 2016). In this view, fiscal consolidation has strong output costs, especially when it is implemented in times of economic slack, but the multiplier associated with stimulus is substantially below 1 regardless of the state of the cycle. If true, this would greatly weaken the case for fiscal policy to play a more important stabilization role in the next downturn. The evidence we present on fiscal expansions for a panel of 17 advanced economies over the last four decades, however, upends this view. We find that large fiscal stimulus packages have had strong and persistent expansionary effects on economic activity. Consistent with the multipliers found in the austerity literature (IMF 2010; Guajardo, Leigh, and Pescatori 2014; Alesina, Favero, and Giavazzi 2019), we do not find that the multiplier associated with stimulus is substantially below 1. We also find that the effects of stimulus are largest in slumps and smallest when the economy is operating close to its output potential. If anything, the current environment of low interest rates further strengthens the case for fiscal stimulus.

We obtain these results by constructing a new dataset of large fiscal stimulus episodes, the Large Fiscal Expansions Database (LFED). To identify possible years of large fiscal expansions, we follow Alberto Alesina and Roberto Perotti (1995) and Alberto Alesina and Silvia Ardagna (2010) and isolate years when the cyclically adjusted budget balance decreased by a large amount. To verify that this signal of the direction of fiscal policy is not a figment of the data, to determine policymakers' motivations in implementing these changes, and to measure their sizes, we follow IMF (2010), Jaime Guajardo, Daniel Leigh,

Jérémie Cohen-Setton is a research fellow and Egor Gornostay a research statistician and quality control coordinator at the Peterson Institute for International Economics. Colombe Ladreit is a PhD student at Bocconi University. We are grateful to Adam Posen for suggesting this research project. We thank Pierre Aldama, Alan Auerbach, Olivier Blanchard, Clémence Dumerger, Patrick Honohan, Daniel Leigh, Joe Gagnon, Josh Hausman, Eric Monnet, Maury Obstfeld, Ángel Ubide, Anna Stansbury, Jeromin Zettelmeyer, and seminar participants at Bank of Japan, Banque de France, France Stratégie, Haverford College, the IMF Fiscal Affairs Department, and PIIE. We are also grateful to Johannes Wieland for making his programs for estimating impulse response functions available online.

and Andrea Pescatori (2014) and Alberto Alesina, Carlo Favero, and Francesco Giavazzi (2019) and exploit contemporary documents produced by the Organization for Economic Cooperation and Development and the International Monetary Fund.¹

We find that measurement errors associated with the cyclically adjusted budget balance are ubiquitous and often correlated with output declines. In the LFED, around 30 percent of the large declines in the cyclically adjusted balance do not correspond to actual fiscal stimulus actions. In many of these cases, the decline in the cyclically adjusted budget balance reflects an unusually large decrease in government revenues following a decline in income or asset prices, which are not captured by standard cyclical adjustment procedures.

We also find that most fiscal expansions are implemented for countercyclical reasons. While not surprising, this helps explain why there are fewer studies of fiscal stimulus than of fiscal consolidation episodes. It also suggests a reason why fiscal policy may seem to have asymmetric effects: reverse causality is a more severe problem for stimulus than for austerity. By clearly identifying and removing from the estimation sample incorrect and endogenous episodes of fiscal stimulus, our approach removes the bias in previous studies that used the Alesina and Perotti (1995) approach to identifying and measuring shifts in fiscal policy.

The paper is organized as follows. Section 2 presents our hybrid approach to identifying and measuring large fiscal expansions for a panel of OECD countries. Section 3 decomposes the downward bias generated by a pure statistical approach to identifying and measuring fiscal expansions and presents our benchmark results and robustness checks for both average and state-dependent multipliers. Section 4 concludes. To allow researchers to improve our judgment calls for classifying fiscal expansions by motivation and complete our data collection efforts, we provide in an online appendix the quotations and citations from the historical record behind our conclusions. Our hope is that this will also help other researchers expand the coverage of our dataset by, for example, successively decreasing the threshold beyond which a decline in cyclically adjusted balances is considered large.

2 The New Large Fiscal Expansions Database

2.1 Motivation

The key challenge in estimating the causal effects of fiscal expansions is identification. To illustrate this, consider a simple linear model where output growth, expressed by ΔY_t (where Y_t is the natural logarithm of real GDP), is assumed to be a function of fiscal policy changes

¹Documents from the European Commission, national authorities, and narrative histories are also sometimes used.

$$\Delta Y_t = \alpha + \beta \Delta F P_t + u_t \tag{1}$$

Unfortunately, existing measures of fiscal policy changes are crude. Even the measure most often used to assess the fiscal stance, the change in the cyclically adjusted budget balance, fails to remove important nonpolicy factors. In particular, fluctuations in both asset and commodity prices (OECD 2006; IMF 2010; Price and Dang 2011; Guajardo, Leigh, and Pescatori 2014; Yang, Fidrmuc, and Ghosh 2015), unstable revenue and expenditure elasticities (Princen, Mourre, Paternoster, and Isbasoiu 2013; Mertens and Ravn 2014; Riera-Crichton, Vegh, and Vuletin 2016), one-off accounting factors (Joumard, Minegishi, André, Nicq, and Price 2008), and revisions in potential output series (Koske and Pain 2008; Cohen-Setton and Valla 2010; Darvas 2013) have been found to generate changes in the cyclically adjusted budget balance that are unrelated to fiscal policy. To the extent that they are driven by a decrease in output growth, including these declines in the cyclically adjusted balance in the estimation of equation (1) would yield underestimates of the true effects of fiscal policy.

Fiscal policy changes also have numerous motivations. Some tax cuts result from views about the incentive effects of marginal tax rates. Some increases in government expenditures reflect a desire to redress past underinvestments or the pursuit of welfare objectives irrespective of budget constraints. Yet other changes in fiscal policy occur because the economy is faltering, as was the case in 2009 when the G20 pushed for a coordinated fiscal stimulus. Including these efforts to stimulate a weak economy when estimating the effects of fiscal policy on short-run fluctuations would, again, be likely to yield underestimates of the true effects.

Therefore, it is useful to think of the change in the cyclically adjusted balance as being composed of three distinct components:

$$\Delta F P_t = \underbrace{[M_0 + M_1(u_t)]}_{\text{Measurement Error}} + \underbrace{ENDO(u_t)}_{\text{Endogenous}} + \underbrace{EXO_t}_{\text{Exogenous}}$$
(2)

The first component, $M_0 + M_1(u_t)$, corresponds to movements in the cyclically adjusted balance that are not driven by fiscal policy actions but rather reflect measurement error. As discussed, several of these measurement errors are likely to be correlated with factors contemporaneously affecting output growth. They are represented by the term $M_1(u_t)$. Removing measurement error is thus not only important for efficiency, but critical for consistency.³ The second component, $ENDO(u_t)$, corresponds to movements in the cyclically

²For simplicity, this presentation omits the cross-country dimension that is present in the rest of the paper. Other factors may include monetary policy shocks, structural reforms, and other nonpolicy disturbances.

 $^{^{3}}$ According to Mertens and Ravn (2013), even measurement errors of the type M_{0} can create

adjusted balance that reflect fiscal policy changes motivated by the state of the business cycle. The third component, EXO_t , corresponds to fiscal policy changes motivated by considerations other than the state of the business cycle, such as reducing tax distortions, improving infrastructure, or redistributing income.

What this decomposition makes clear is that the assumption of zero contemporaneous correlation between a change in a crude measure of fiscal policy and other factors affecting output growth— $\mathbb{E}(\Delta F P_t u_t) = 0$ —is clearly violated. Hence, unbiased and consistent Ordinary Least Square estimates of the impact of fiscal policy cannot be obtained from equation (1). The approach of this paper consists of cleaning $\Delta F P_t$ of its measurement error and endogenous components and thus exploit only policy changes not motivated by a faltering economy to estimate the effects of fiscal stimulus. With a measure of exogenous fiscal changes, EXO_t , the effects of fiscal expansions on output can be estimated by fitting a simple distributed lag model of the following form:

$$\Delta Y_t = \alpha + \sum_{j=0}^n \beta_j EXO_{t-j} + u_t \tag{3}$$

2.2 Construction of the Large Fiscal Expansions Database

Figure 1 shows our strategy to identify and measure large exogenous fiscal stimulus episodes. First, we searched for possible episodes of large fiscal expansions by using a statistical approach like that of Alesina and Perotti (1995), where large movements in a cyclically adjusted measure of the fiscal balance are used to identify important changes in fiscal policy. For the years when the decline in this measure is large, we then read IMF Article IV reports, OECD Economic Outlooks, and OECD Economic Surveys to see if the description of fiscal policy aligned with the movement in the statistical indicator. Doing so allowed us to identify what Alesina (2010) calls incorrect episodes that are erroneously captured by the statistical approach. When a fiscal action was confirmed by the policy record, we classified it as endogenous if it was taken in response to macroeconomic fluctuations or as exogenous if it was the byproduct of other considerations. For large exogenous fiscal stimulus episodes, we collected real-time narrative measures of the fiscal policy.

2.2.1 Identifying Possible Large Fiscal Expansions

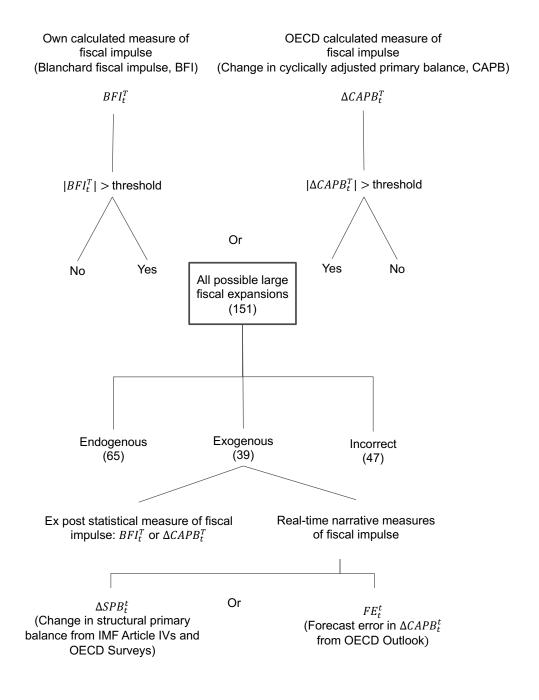
This paper looks for indications of large fiscal changes in an unbalanced panel of 17 countries from 1960 to 2006.⁴ The annual data are collected from the OECD *Economic Outlook* N°100 (June 2016), except when older issues of the *Economic Outlook* provide longer time series.⁵

attenuation bias.

⁴Our sample covers the same countries as Guajardo, Leigh, and Pescatori (2014).

⁵This is the case for Australia, Austria, France, Germany and Ireland. See Appendix A for details on data sources.

Figure 1: Identifying and Measuring Shifts in Fiscal Policy: A Guide



Two approaches are generally used to obtain a cyclically adjusted measure of the fiscal balance for a panel of countries. The most straightforward approach is to use a measure already calculated by an international organization such as the IMF or the OECD, a measure we refer to as the cyclically adjusted primary balance (CAPB). A change in the latter is denoted as $\Delta CAPB_t^T$, where the subscript specifies the year t when the change occurs, and the superscript points to the vintage of the estimate. In the statistical approach, ex post data of vintage T are generally used. A disadvantage of this approach is that such measures are available only since the mid-1980s or early 1990s for most countries. Thus, another approach is often used building on Olivier Blanchard (1990, p.12)'s suggestion to calculate a cyclically adjusted measure of the fiscal balance as 'the value of the primary surplus which would have prevailed, were the unemployment at the same value as in the previous year, minus the value of the primary surplus in the previous year.' In what follows, a change in this measure is referred to as the Blanchard fiscal impulse and is denoted as BFI or BFI_t^T .

To implement Blanchard (1990)'s idea, we closely follow Alesina and Perotti (1995). For each country, we first estimate the relationship between certain components of government revenues and expenditures (respectively R_t and G_t) expressed in percentages of GDP and the unemployment rate (U_t) .⁶ The estimated coefficients together with the previous year's unemployment rate (U_{t-1}) are then used to calculate primary expenditures (G_t^*) and revenues (R_t^*) adjusted for changes in the unemployment rate.⁷ The BFI is then calculated as the difference between the primary balance adjusted for changes in unemployment in period t and the actual primary balance in period t-1. A negative BFI means that the government spent more or levied less in taxes than what the state of the economy would have normally implied, suggesting a possible expansionary fiscal stance.

$$G_t = \phi_0 + \phi_1 Trend + \phi_2 U_t + \epsilon_t \tag{4}$$

$$R_t = \gamma_0 + \gamma_1 Trend + \gamma_2 U_t + \eta_t \tag{5}$$

$$G_t^* = \hat{\phi}_0 + \hat{\phi}_1 Trend + \hat{\phi}_2 U_{t-1} \tag{6}$$

$$R_t^* = \hat{\gamma}_0 + \hat{\gamma}_1 Trend + \hat{\gamma}_2 U_{t-1} \tag{7}$$

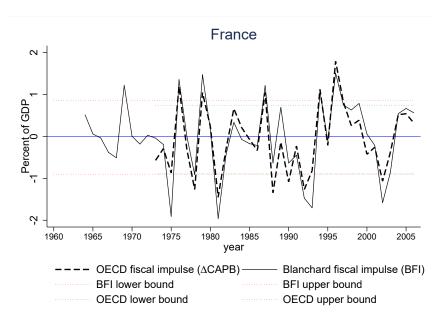
$$BFI_t^T = [R_t^* - G_t^*] - [R_{t-1} - G_{t-1}]$$
(8)

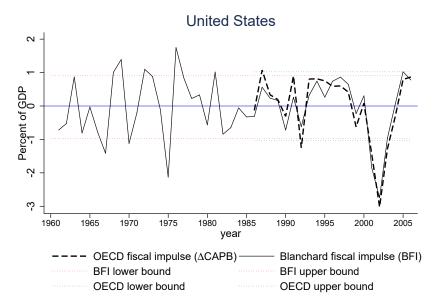
Figure 2 compares the BFI obtained for France and the United States with $\Delta CAPB_t^T$ as calculated by the OECD. Clearly, the BFI approach is useful in extending the time coverage as it adds 10 and 20 more years of observations for France and the United States, respectively. While the two lines move closely together, the exact magnitudes of the two indicators are not always the same.

 $^{^6}$ For government expenditures, only transfers are adjusted. This follows Alesina and Perotti (1995) and the approach used by the OECD (Girouard and André 2005) to cyclically adjust expenditures.

⁷Other components are then added without any adjustment.

Figure 2: Comparison between OECD's $\Delta CAPB_t^T$ and own calculated BFI_t^T





Note: CAPB = cyclically adjusted primary balance. BFI = Blanchard fiscal impulse. The lower and upper bounds are for 1-year episodes, meaning that they correspond to the mean +/- one standard deviation. Our 2-year thresholds (mean +/- 1.5*s.d.) are not represented here but can be found in appendix table B.1.

Focusing on large shocks requires the definition of thresholds. In their seminal paper, Alesina and Perotti (1995) use a fixed threshold of 1.5 percent of GDP for all countries. Given the heterogeneity across countries in the mean (μ_i) and standard deviation (σ_i) of changes in fiscal policy documented in appendix table B.1, the use of country-specific thresholds is preferred. More specifically, a decline in the cyclically adjusted fiscal measures of country i is considered large if it is larger in absolute value than $\mu_i - \sigma_i$ in a single year or $\mu_i - 1.5\sigma_i$ over two years (with a size of at least $\mu_i - 0.5\sigma_i$ in each year). The second criterion helps capture episodes that are large but happen over several years.

2.2.2 Eliminating False Positives and Classifying by Motivations

Out of our sample of 691 BFI_t^T and 436 $\Delta CAPB_t^T$ observations, we obtain 151 country-year pairs of large declines (appendix table B.2). For each of these, we read contemporaneous policy documents (typically OECD Economic~Outlooks, OECD Economic~Surveys, and IMF Article~IV~Reports for the previous, current, and following years) to assess whether the declines in these statistical measures correspond to fiscal actions and document their rationale.⁸

In our sample, we find that the standard statistical measure is misleading about 30 percent of the time. For these country-year pairs, we are able to find specific economic or budgetary developments that cause the standard statistical measure to inaccurately identify the size of an episode. Appendix table B.3 strikes through the incorrect episodes from the list of all possible episodes. Clearly, the problem is widespread. All countries, except Japan, display years that are wrongly identified as having large fiscal expansions by the standard statistical approach. Additional documentation on each of these cases is available in a companion appendix provided as supplementary material, in which we provide the citations for each data point that lead us to classify an episode as incorrect.

In some cases, we can see directly that a decline in a cyclically adjusted fiscal measure was a misleading indicator of actual fiscal actions from descriptions provided in the policy records, which reveal either a lack of government intent or a desire to implement a contractionary rather than expansionary fiscal stance. In France in 1993, "the sharp widening in the...deficit [is described as] largely unintended." In Portugal in 1978, "fiscal policy [is described as] not intended to give impetus" by the OECD. Rather than being expansionary, the fiscal stance in Spain is described as "mov[ing] towards restriction" in the early 1990s. Similarly, the OECD emphasizes the "strongly restrictive stance to monetary and fiscal policy" in Italy in 1981. It also points out that, "dictated by balance-of-payments considerations, fiscal policy was tightened" in Sweden in 1980.

⁸When needed, we complement these sources with other IMF *Staff Report* such as *Recent Economic Developments*. We also sometimes rely on publications by the European Commission and check our findings against narrative histories when available.

⁹46 country-year pairs out of 152.

In a few cases, declines in the cyclically adjusted fiscal measures arise because of one-off accounting events. For instance, a "Fund for Railway Infrastructure, which owns the railway infrastructure, was created" in Belgium in 2005. As the general government became "liable for [its] corresponding debt of 7.4 billion euros...government debt increase[d] by 2.5 percent of GDP." While these one-off transfers onto the balance sheet of the general government are unlikely to be correlated with the state of the economy, such measurement errors do not just create noise and decrease the efficiency of statistical estimates, but also bias the estimated impact of fiscal policy toward zero.

More importantly, the bulk of incorrect episodes identified by the statistical approach arise because of falls in revenue elasticities and increases in spending elasticities. In Belgium in 1988, it is, for example, argued that the backsliding in fiscal finances since 1987 "result[s] essentially from the unintended fall in the apparent elasticity of revenue with respect to GDP observed during this period." Similarly, the "large overshoot" appears mostly "due to [the] operation of stabilizers" rather than discretionary fiscal action. That such a problem arises is not surprising given the evidence documented in IMF (2010) and Guajardo, Leigh, and Pescatori (2014) for fiscal consolidation episodes and the well-understood problem with nonconstant revenue and expenditure elasticities (Princen, Mourre, Paternoster, and Isbasoiu 2013). But the problem is more widespread than generally understood.

As in Christina Romer and David Romer (2010), we then separate correct episodes of large fiscal expansions into two broad categories: those that happened in response to factors likely to affect output growth in the near future (endogenous) and those taken for other reasons (exogenous). If policy documents indicate that some of the measures contributing to the large fiscal expansion were implemented because of a desire to respond to current or prospective economic conditions, we classify the entire episode as endogenous. This is more restrictive than what is usually done in the narrative literature, which allocate the total change in fiscal policy between exogenous and endogenous motivations.

Examples of endogenous episodes include Belgium in 1972, for which the OECD stated that "the slowdown in economic growth...led the Belgian authorities to modify the posture of economic policy in a more expansionary direction"; Canada in 1983, when the government's "Special Recovery Program...to stimulate private sector investment, formed the centerpiece of the budget"; or the United States in 1970 when the "policy was eased in the first half of 1970 to limit the downturn of the economy." The online appendix provides similar quotes for the other episodes that we consider endogenous.

This leaves us with a set of exogenous episodes, which can be used for causal inference. Examples of episodes that are considered exogenous are the United States in 1982–83 and 2002–2003 and France in 2002, when tax cuts were implemented to increase long-term growth; Ireland in 1978, when the government wanted to reduce hysteresis unemployment through excess demand; and the United Kingdom in 2001–2003 when the government wanted to redress past underinvestment. Detailed documentation on each fiscal policy change is in

the online appendix, in which we provide citations to show how we determine the motivation.

2.2.3 Measuring Large Fiscal Expansions

To use the common metric of fiscal multipliers, the size of the fiscal impulse needs to be documented. Two approaches are usually followed. First, in the statistical literature exemplified by Alesina and Ardagna (2010), BFI_t^T (where subscript t refers to the time period of the fiscal shock and superscript T refers to vintage of the data series) is used not only to identify shifts in fiscal policy, but also to quantify them. As BFI_t^T was revealed to often be a misleading indicator of the direction of a shift in fiscal policy, there is reason to believe that it might also be a misleading indicator of its magnitude, even when the direction is accurate. Second, in the narrative approach exemplified by Romer and Romer (2010), the fiscal impulse is generally measured as the legislative forecasts of the expected cumulative effect on tax revenues and government expenditures.

In this paper, we propose two alternative measures. The first one is the change in structural primary balance (denoted ΔSPB_t^t) as estimated in real time and published in IMF Article IV Reports and OECD Economic Surveys. An advantage of these estimates is that they have "reality checks," as the OECD and IMF exploit their local presence in the member countries and hold extensive discussions. The OECD Economic Surveys are, for example, reviewed by representatives of OECD member state governments, gathered by the Economic and Development Review Committee (EDRC), including the country under review. A disadvantage of this measure, however, is that the underlying method used to obtain these estimates is not uniform across countries, time periods, and sources. We partly address this concern in section 3.5 when introducing the Local Projection–Instrumental Variable framework.

Another disadvantage is that the change in structural primary balance ignores that private agents may respond to fiscal policy before the shift in policy is actually implemented (Blanchard and Perotti 2002; Mertens and Ravn 2012; Ramey 2016). The preannounced 2014 Japanese consumption tax increase illustrates this problem: consumers brought forward their purchases of durable goods in the quarter before the implementation of the new tax. In this case, relying on the implementation date would generate a spuriously big fiscal multiplier. To address this issue, we construct forecast error measures (FE) for the change of the cyclically adjusted primary balance from multiple editions of the OECD Economic Outlook. Like Auerbach and Gorodnichenko (2013) for government expenditures, Abdul Abiad, Davide Furceri, and Petia Topalova (2016) and Davide Furceri and Bin Grace Li (2017) for public investment, and Patrick Blagrave, Giang Ho, Ksenia Koloskova, and Esteban Vesperoni (2017) for tax revenues, we calculate forecast errors as the difference between the forecast and realized values, thereby purging fiscal variables of their predictable compo-

nents.¹⁰ Another advantage is that fiscal forecasts from the OECD Economic~Outlook are calculated using a uniform method for all countries, in contrast to the estimates of ΔSPB obtained from IMF Article~IV~Reports and OECD Economic~Surveys.

OECD fiscal forecasts, however, start only in the 1980s and do not eliminate the impact of one-off operations (Journard, Minegishi, André, Nicq, and Price 2008). In fact, it is only since the December 2008 $Economic\ Outlook$ that the OECD introduced such systematic correction with the creation of an underlying fiscal balance measure (NLGQU), which removes net one-offs (NOOQ) from the cyclically adjusted net lending variable (NLGQA). For both reasons, we use ΔSPB_t^t as our baseline measure rather than FE_t^t .

2.2.4 Limitations

A first limitation of our approach is its focus on large fiscal expansions. By construction, small fiscal stimulus cannot be identified with our approach. This could be a problem if the incremental impact of a fiscal expansion on economic activity is related to the size of the package. While we cannot exclude this possibility for fiscal expansions, it has not been documented for fiscal consolidations (IMF 2010). Both small and large fiscal consolidations have the same multiplier.¹² The LP-IV approach introduced in section 3.5 helps address this issue.

A second limitation is that our approach may fail to pick up episodes of fiscal expansions that did not translate into a sufficiently large decline in BFI_t^T or in $\Delta CAPB_t^T$. This could happen either because of one-off factors that increased fiscal balances at the time of the stimulus or because the fiscal expansion occurred concurrently with asset or commodity price booms that boosted government revenues. In these cases, there might be no change in the observed structural deficit, but there is a fiscal expansion. From the perspective of getting consistent estimates, missing large intended fiscal expansions does not constitute in itself a problem. If these missing expansions were endogenous, they should not be in the regression. If they were exogenous, missing them is unfortunate as it makes our database incomplete, but this only adds noise to our regression results.

A third limitation has to do with determining the intent of policymakers for some fiscal expansions from narrative evidence. While it is, for example, relatively clear-cut that an absence of new tax measures can be equated with a neutral tax stance, defining a neutral

¹⁰We compare the forecast made in year t-1 for the change in fiscal policy in year t, $\Delta CAPB_t^{t-1}$, to an average of estimates of the fiscal policy change that occurred in year t according to the following five issues of the OECD *Economic Outlook*. Thus, $FE_t = \Delta CAPB_t^{t-1} - \sum_{j=0}^4 \Delta CAPB_t^{t+j}/5$. This choice helps limit the problem of revised ex post estimates.

¹¹A telling example is Sweden in 2001, when one-off factors made the OECD *Economic Outlook* the post 2002 vintages estimates of the 2001 fiscal change look contractionary despite clear evidence of a fiscal stimulus.

¹²In IMF (2010), figure 3.2 shows the effect for all consolidations, while figure 3.10 shows the effect of large fiscal consolidations, defined as discretionary deficit cuts larger than 1.5 percent of GDP.

stance for expenditures (Carnot and Castro 2015) is harder and sometimes means that one has to judge whether an expenditure slippage was intentional or not. Missing unannounced but intentional expenditure slippages could bias our results. Similarly, if policymakers postpone tax cuts and expenditure increases until the economy strengthens (beyond the normal economic dynamic that we control for) then the fiscal expansions might still be associated with business cycle developments. This would bias our results toward associating economic overheating with fiscal expansions and overestimating expansionary effects.

A fourth limitation concerns the effect of fiscal policy announcements. As argued by Alesina, Favero, and Giavazzi (2015), fiscal policy in year t generally consists of three components: unexpected shifts in fiscal variables (announced upon implementation in year t), shifts implemented at time t but announced in previous years, and future announced changes (announced at time t for implementation in some future year). While our approach allows us to measure these first two components, it does not capture the third one.

2.3 Properties of the New Database

Out of 151 country-year pairs of large declines in the cyclically adjusted fiscal measures, 104 are found to reflect government actions. The other 47 correspond to large declines in cyclically adjusted fiscal balances that happen despite an absence of fiscal stimulus. Among 104 LFE identified, two-thirds (65) are found to be motivated by countercyclical reasons and one-third (39) are not.¹³

The 104 endogenous and exogenous LFE are shown in the diagrams in Figure 3. An interesting feature is that fiscal policy changes appear synchronized across OECD countries. That this would be the case against the backdrop of the synchronized recessions of the early 1970s, 1980s, and 1990s is hardly surprising. But even so-called exogenous fiscal changes appear correlated across countries, with many implementing, for example, tax cuts to improve potential growth in the early 2000s.

As explained in section 2.2.1, some episodes are identified and measured in the statistical approach using BFI_t^T while others are identified and measured using the official measure calculated by the OECD, $\Delta CAPB_t^T$. Since both are similar conceptually and calculated using ex post data, in what follows that statistical impulse is referred to as BFI irrespective of whether it is actually BFI_t^T or $\Delta CAPB_t^T$. This avoids carrying both notations.

Figure 3 displays different measures of the fiscal impulse. For all episodes, the BFI measure is displayed. For exogenous episodes, the ΔSPB and FE measures are also displayed

¹³Note that Australia 1965 is discarded from our original list of 39 exogenous country-year pairs. Australia 1965 and Australia 1966 are picked up by our identification method, with Australia 1966 identified as an endogenous episode. Once this is taken into account, Australia 1965 no longer meets our 1-year threshold. For that reason, we do not keep it. This is the only country-year pair for which this issue arises. For all other consecutive exogenous-endogenous cases, the exogenous episode still meets the threshold once the endogenous year is dropped.

when available. The BFI is generally larger in absolute value than the other measures. In some cases, this clearly reflects measurement errors. According to the BFI measure, Germany implemented 4 percent of GDP fiscal stimulus in 2001, which is clearly incorrect. According to ΔSPB , the stimulus amounted to only around 1 percent of GDP. As described in our online appendix, this impulse was widely expected as part of the 2000 Tax Relief Act. It is thus not surprising that the FE measure is actually positive.

Our key identifying assumption is that our constructed series of large fiscal expansions is, indeed, exogenous. While the contemporaneous exogeneity of the constructed series cannot be tested, it is possible to test whether the series is predictable on the basis of past information. To test this hypothesis, a multinomial response model (as in Cloyne (2013), Jordà and Taylor (2016), and Mertens and Ravn (2013)) and Granger causality tests (as in Cloyne (2013)) are estimated. As in the rest of the literature, lags of the growth rate of real GDP, inflation, short-term interest rate changes, and the government debt-to-GDP ratio are included as regressors for both tests. Unobserved heterogeneity and common trends are also controlled by including country and year fixed effects. Since a lag of the growth rate of real GDP is controlled for in our baseline model specification for output impulse responses (see section 3.1), it is not included in the joint tests in table 1.

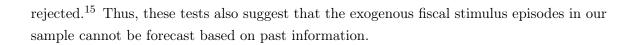
Table 1: Predictability tests for exogenous episodes

Test	Test statistic	p-value
Logit (Likelihood ratio test)	4.30	0.231
Granger causality (F-test) for ΔSPB	1.36	0.267
Granger causality (F-test) for FE	0.91	0.443

Note: In both models, regressors comprise one lag of each of the macroeconomic variables as well as one lag of the dependent variable. These results are robust to using two or three lags. Both models include country and year fixed effects. To correct for possible misbehavior of standard errors, statistical inference from the Granger causality test is based on Driscoll and Kraay (1998) standard errors.

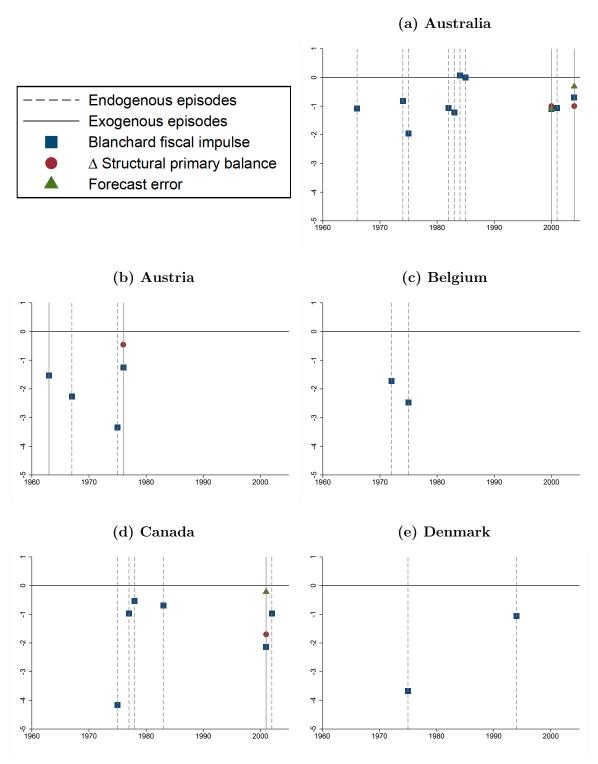
The top line of table 1 displays the test statistic and p-value associated with the null hypothesis that these variables have no explanatory power in a conditional logit model, where the dependent variable takes a value of 1 when a large fiscal stimulus happens. With a p-value of 0.231, the hypothesis that the exogenous episodes are not predictable cannot be rejected. The next two lines of table 1 show the result of the single-equation version of the Granger causality test, which exploits information about both the timing and the size of an exogenous fiscal stimulus, with either ΔSPB or FE used to measure the size of the fiscal impulse. For both measures, the non predictability hypothesis cannot be

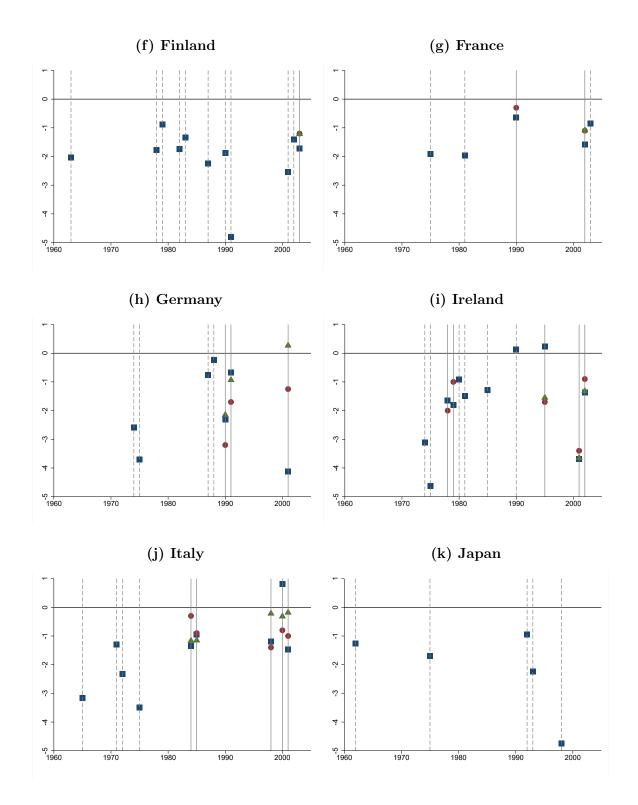
¹⁴Òscar Jordà and Alan Taylor (2016) and Mertens and Ravn (2013) used the unconditional probit estimator, which in panel data with fixed effects suffers from the incidental parameters problem, leading to inconsistent parameter estimates. We thus use the conditional logit estimator instead.

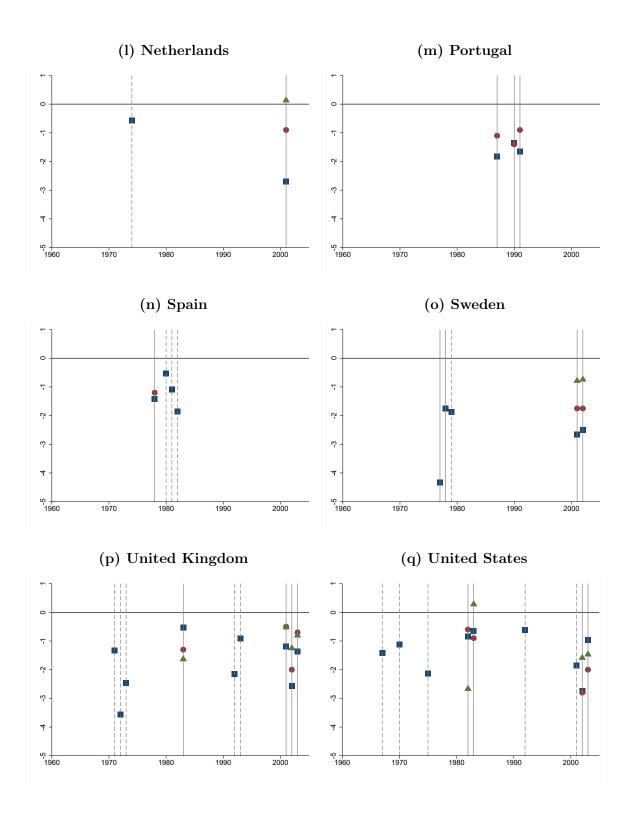


The difference in results between the two measures is explained mostly by the fact FE has fewer observations than ΔSPB .

Figure 3: Endogenous and exogenous fiscal expansions







3 Effects of Fiscal Expansion on Economic Activity

3.1 Baseline Specification

Following Jordà (2005), we use the method of local projections (LPs) to estimate impulse response functions (IRFs) of output and other macro aggregates. The specification we consider takes the following form:

$$Y_{i,t+h} - Y_{i,t-1} = \alpha_i^h + \lambda_t^h + \phi^h \Delta Y_{i,t-1} + \sum_{j=0}^n \beta_j^h \Delta F P_{i,t-j}^{a,b,c,d,e} + u_{i,t+h}$$
 (9)

where subscript i indexes countries, subscript t indexes years, Y is the logarithm of real GDP (or hours, real wages...) or the level of the unemployment rate (or interest rates...), α_i denotes country-fixed effects, λ_t denotes year-fixed effects, and $u_{i,t}$ is an error term. The variable of interest ΔFP has a number of different superscripts because we estimate equation (9) based on several measures of fiscal impulse.

The β_j coefficients are the contemporaneous and lagged effects of fiscal expansions. The ϕ coefficient is an autoregressive term. Only one lag of the dependent variable and one lag of the fiscal impulse are chosen as this combination maximizes the within R^2 statistic in output on ΔSPB regressions. Including a lag of the fiscal impulse in the baseline model allows us to capture a possible serial correlation in it. In contrast with the traditional vector autoregression (VAR) method, which uses the one-period-ahead expectation to form the two-period-ahead expectation, LPs use a separate regression for each forecast horizon h. Hence, equation (9) is reestimated for each IRF horizon h = 1, 2, 3, 4 (with h = 1 indicating the initial year of a fiscal expansion episode), and β_0^h is stored after each such regression.

To consistently estimate β using the standard fixed effects estimator, strict exogeneity of explanatory variables conditional on the unobserved effect —i.e., $\mathbb{E}(u_{i,t}|\Delta FP_{i,1},...,\Delta FP_{i,T},\lambda_1,...,\lambda_T,\alpha_i)=0$ for all t=1,...,T —needs to be satisfied. This assumption implies that fiscal policy changes in each time period are uncorrelated with the idiosyncratic error in each time period, i.e., $\mathbb{E}(\Delta FP_{i,s}u_{i,t})=0$ for all s,t=1,...,T. This is a stronger assumption than just assuming zero contemporaneous correlation in the single-country model of section 2.1 for OLS consistency.

As the most obvious concern when regressing output growth on a fiscal shock is that the result does not distinguish between the effect of the shock and that of normal output dynamics (Romer and Romer 2010), our benchmark specification includes one lag of output growth as a regressor. ¹⁶ Including lagged values of the dependent variable as controls, however, makes the fixed-effect estimator inconsistent because of the violation of the strict exogeneity assumption (Nickell 1981) and is likely to generate an upward bias. ¹⁷ The in-

¹⁶The same argument holds for other outcome variables.

¹⁷With positive coefficients on the lagged dependent variable and on (the negative of) the fiscal

consistency of the estimated coefficients is, however, unlikely to be sizable as it is of order 1/T. In fact, we verify this by reestimating equation (9) using the Arellano and Bond (1991) generalized method of moments estimator that is designed to address this problem. Like Guajardo, Leigh, and Pescatori (2014) for fiscal consolidations, we find a minor difference in results between the fixed-effect and Arellano and Bond (1991) estimators and so use the fixed-effect estimator in the rest of the paper.

Despite the inclusion of time-fixed effects, the standard Pesaran (2004) and Frees (1995) statistical tests reject the null hypothesis that the residuals from the fixed-effect estimation are uncorrelated across countries. Using commonly applied robust standard errors (e.g. White, clustered [Rogers], or Newey-West) would thus be inappropriate for statistical inference as those techniques correct only for heteroscedasticity and serial correlation within countries (or clusters of countries). Instead, we use Driscoll and Kraay (1998) standard errors, which are designed to address the problem of cross-sectional correlation in addition to within-country heteroscedasticity and serial correlation.¹⁸

3.2 Biases from Incorrect and Endogenous Fiscal Expansions

When the set of episodes considered include all large declines in BFI_t^T or in $\Delta CAPB_t^T$, the variable of interest is called $\Delta FP_{i,t}^a$ and takes the value BFI_t^T or $\Delta CAPB_t^T$ for large declines and zero otherwise. When narrative evidence is used to identify instances of large declines in BFI that do not reflect actual fiscal actions, the variable of interest is called $\Delta FP_{i,t}^b$. For correct episodes, it takes the value BFI. For incorrect episodes, $\Delta FP_{i,t}^b$ is equal to zero. When attention is further restricted to correct episodes that were not motivated by countercyclical reasons, the variable of interest is called $\Delta FP_{i,t}^c$. It takes the value BFI for exogenous episodes, and zero otherwise.

To assess whether the BFI measure might itself generate bias, two alternative narrative fiscal impulse measures obtained from OECD and IMF reports are used. We let $\Delta FP_{i,t}^d = \Delta SPB_t^t$ for exogenous episodes and zero otherwise, where the superscript t shows that this is a real-time rather than ex post measure. We let $\Delta FP_{i,t}^e = FE_t^t$ for the subset of exogenous episodes for which this variable is available and zero otherwise.

Figure 4a compares the relative path of the log of real GDP in response to a 1 percent of GDP fiscal expansion based on the different definitions and measurements of the fiscal shock. The difference between $\Delta F P_{i,t}^a$ and $\Delta F P_{i,t}^b$ shows how the inclusion of episodes that are incorrectly identified (i.e., that do not reflect a fiscal action by the government) can bias

impulse, we expect the fixed-effect estimator to overestimate the impact of fiscal expansions.

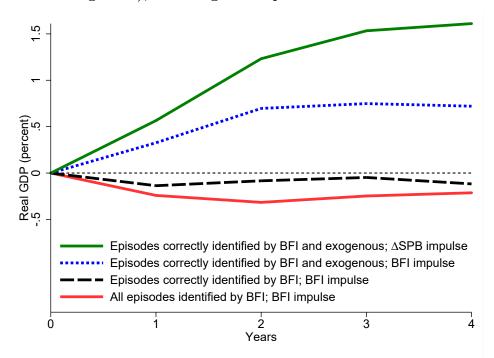
¹⁸The asymptotic properties of this variance-covariance matrix estimator do not rely on the number of countries, N, but rather on the number of time periods, T. As with Newey-West standard errors, the number of lags up to which the residuals may be autocorrelated needs to be specified. We follow the simple rule of thumb of the xtscc Stata procedure for selecting the number of lags, m(T), as $floor(\frac{4(T/100)^2}{9})$. In our setup, this leads to m(T) = 3.

the results. The difference between $\Delta F P_{i,t}^b$ and $\Delta F P_{i,t}^c$ reflects the bias that can arise from mixing countercyclical and exogenous episodes. The difference between $\Delta F P_{i,t}^c$ and $\Delta F P_{i,t}^d$ measures the bias due to an incorrect measurement of the fiscal impulse associated with a correctly identified exogenous fiscal expansion.

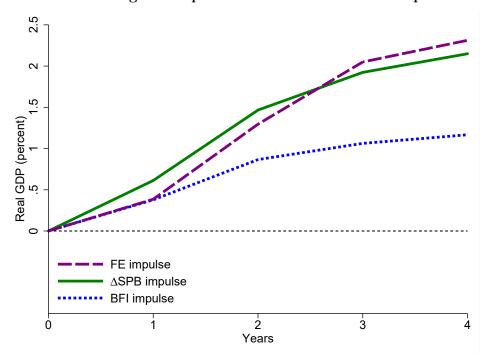
The results are striking. First, they illustrate that relying only on the statistical approach to both identify and measure large fiscal expansions generates the curious finding that fiscal expansions are contractionary. In other words, the result of Alesina and Ardagna (2010) is driven not only by expansionary fiscal consolidations, but also by contractionary fiscal expansions. As in IMF (2010) and Guajardo, Leigh, and Pescatori (2014), it arises because of measurement errors —due mostly to movements in revenue and expenditure elasticities—that are correlated with movements in economic output. Second, the results illustrated in figure 4a emphasize the importance of classifying fiscal expansions by motivation. The blue dotted line shows the path of real GDP following large fiscal expansions correctly identified and primarily motivated by exogenous factors, while the black dashed line shows the effects of all correctly identified episodes (i.e., both exogenous and endogenous fiscal expansions). It reveals that the downward bias created by mixing these two kinds of fiscal expansions is particularly large.

Figure 4: Identification and measurement pitfalls

(a) Identification: All identified by statistical approach, correctly identified (endogenous and exogenous), and exogenous episodes



(b) Measurement: Exogenous episodes with different fiscal impulse measures



Note: BFI = Blanchard fiscal impulse, ΔSPB = Change in the structural primary balance, FE = forecast error. The lines indicate the cumulative percentage change in real GDP years 1 to 4 relative to year 0 in response to a fiscal shock of 1 percentage point of GDP in year 1. In the bottom panel, only exogenous episodes for which the FE measure exists are used. This explains the difference between the solid green lines and the dotted blue lines in the two panels.

3.3 Biases from Mismeasurement of Fiscal Impulse

The top two lines of figure 4a illustrate the difference in results that arises from using either BFI or ΔSPB as a policy impulse. The fact that the IRF associated with ΔSPB is higher than that associated with BFI illustrates that the fiscal impulse obtained from IMF and OECD reports is typically smaller than that estimated with ex post data (BFI_t^T) and (BFI_t^T) . This difference is due to measurement errors in either the statistical impulse or measurement errors in the policy records impulse.

Differences in follow-up could explain measurement errors in the policy records impulse. If fiscal plans rather than fiscal actions are recovered from the policy records, and if governments tend to deviate from their fiscal plans (i.e. spend more or tax less than planned), then the policy records impulse would underestimate the actual size of fiscal expansions.¹⁹ This explanation is, however, not very convincing here. In fact, rather than collecting fiscal plans as in IMF (2010), Guajardo, Leigh, and Pescatori (2014), and Alesina, Barbiero, Favero, Giavazzi, and Paradisi (2017), we collect fiscal actions as documented in the policy record in years following the fiscal impulse. For instance, the ΔSPB fiscal impulse for Australia in 2000 is obtained from documents published in 2003 rather than from documents published before the implementation of the policy change.

To compare the impact of different measure of the fiscal impulse, figure 4b reproduces the results of our baseline specification for the sub-sample for country-year pairs where the forecast errors fiscal impulse measure is also available. Two things are worth noting. First, IRFs for both BFI and ΔSPB_t^t are slightly higher in figure 4b than in figure 4a. This suggests that the fiscal expansions that took place in the earlier part of our sample period and for which we lack data on forecast errors —Austria, Ireland, and Spain in the late 1970s, and Portugal in the late 1980s and early 1990s —were less effective at stimulating output than those that happened in more recent years. Second, the unanticipated parts of fiscal policy changes appear to not have stronger effects than the part that is anticipated. This finding contrasts with those obtained by Mertens and Ravn (2014) and Ramey (2019) with quarterly data.

3.4 Benchmark Effects

We start by reporting in figure 5 the IRF to a 1 percent of GDP fiscal expansion for (the natural logarithm of) real GDP and the unemployment rate. For all IRFs, the shock corresponds to fiscal impulse measure ΔSPB that happens in year 1, and the forecast horizon is

¹⁹Roel Beetsma, Oana Furtuna, and Massimo Giuliodori (2017) investigate this hypothesis for fiscal consolidations and whether differences in follow-up between expenditure-based and tax-based fiscal plans can explain the difference in tax and spending multipliers.

²⁰Forecast errors cannot be constructed because of data unavailability for Austria (1976), Ireland (1978), Ireland (1979), Portugal (1987), Portugal (1990), Portugal (1991), and Spain (1978).

4 years. A large fiscal stimulus is found to set off a major and persistent expansion in the economy.

According to our estimates, a 1 percent of GDP fiscal expansion is associated with a 1.5 percent peak cumulative increase in GDP. The increase in output is statistically significant in the first year of the shock and builds over time to reach its peak after two years. The size of the effect is the same (in absolute value) as that obtained by Jordà and Taylor (2016) for fiscal consolidations: they report that real GDP is pushed down on average by over 0.57 percent each year for every 1 percent in fiscal consolidation. In this sense, our results do not support the hypothesis that the contractionary multiplier is bigger (in absolute value) than the expansionary multiplier (Barnichon and Matthes 2016).²¹

The time profile of the decrease in the unemployment rate is similar to that of the increase in output. A 1 percent of GDP fiscal stimulus is associated with a maximum cumulative decrease in the unemployment rate of 0.4 percentage points after three years. The effect is almost the same as that obtained by Guajardo, Leigh, and Pescatori (2014), who find a 0.3 cumulative increase in the unemployment rate two years after the start of a 1 percent of GDP fiscal consolidation. The fact that on unemployment but not on GDP we obtain the same results as Guajardo, Leigh, and Pescatori (2014), for fiscal consolidation, suggests either that their results are smaller due to specification choices or that the link between output growth and unemployment is different during fiscal stimulus and fiscal consolidations. As Jordà and Taylor (2016) do not report results for unemployment, further work is needed to distinguish between these hypotheses.

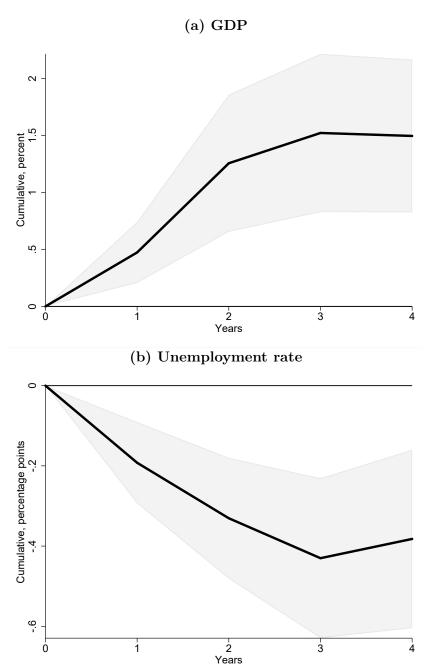
The top two panels of figure 6 summarize the results of reestimating our baseline specification for the contributions to real GDP of final private domestic expenditures and net exports. The contribution of real final private domestic demand is defined as $\frac{FPDV_{t-1}}{GDPV_{t-1}} \times g_{FPDV,t}$, where FDPV and GDPV respectively denote real final private domestic expenditures and real GDP, and $g_{FPDV,t}$ denotes the growth rate of FPDV.²² Similarly, the contribution of net exports is defined as $\frac{NXV_{t-1}}{GDPV_{t-1}} \times g_{NXV,t}$, where NXV and GDPV respectively denote real net exports and real GDP, and $g_{NXV,t}$ denotes the growth rate of real net exports.²³ We find that a decrease in net exports partly offsets expansionary effects on total private domestic demand (figure 6b). According to our estimates, net exports reduce GDP growth

²¹Guajardo, Leigh, and Pescatori (2014) obtain smaller effects on real GDP (between 0.5 and 0.8 percent of GDP after two years depending on the specification) than Jordà and Taylor (2016) despite using the same IMF narrative database of fiscal consolidations. If true, our results together with those of Guajardo, Leigh, and Pescatori (2014) would suggest asymmetric effects of fiscal policy, but not in the direction argued by Barnichon and Matthes (2016).

 $^{^{22}}$ We compute real final private domestic (FPDV) expenditures as FFDV-CGV-IGV, where FFDV is real final domestic expenditures, CGV is real government final consumption expenditures, and IGV is real government investment expenditures in the OECD $Economic\ Outlook\ Database$. For Italy, Germany, Portugal, and Spain, code IGV is not consistently available. FPDV is thus simply equal to FFDV-CGV.

 $^{^{23}\}mathrm{We}$ compute real net exports as GDPV minus FFDV as in Alesina, Favero, and Giavazzi (2019).

Figure 5: Impact of 1 percent of GDP fiscal expansion: GDP and unemployment rate



Note for panel (a): The solid line indicates the cumulative percentage change in real GDP in years 1 to 4 relative to year 0 in response to a fiscal shock of 1 percentage point of GDP in year 1. Note for panel (b): The solid line indicates the cumulative percentage point change in the unemployment rate. Note for panels (a) and (b): The shaded areas represent one standard–error bands.

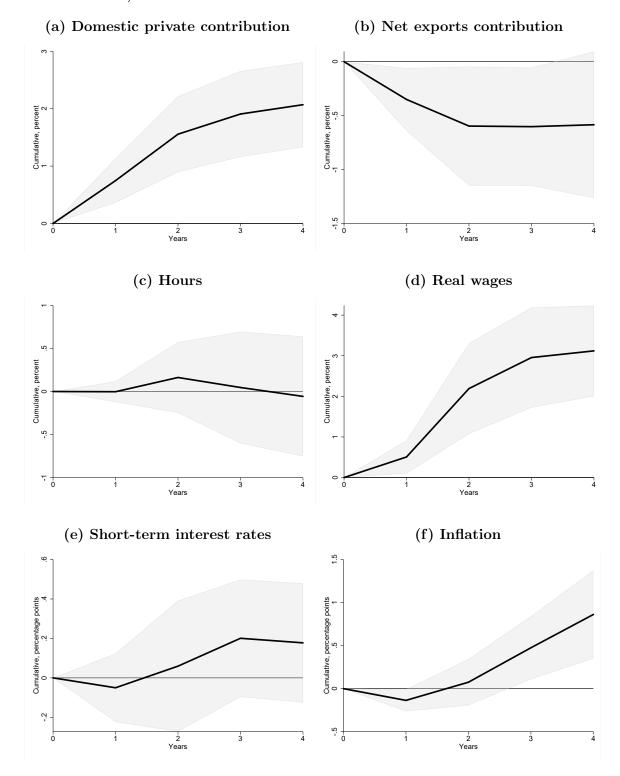
by up to a cumulative 0.5–0.6 percentage points.²⁴

Despite the standard prediction of theoretical dynamic stochastic general equilibrium models that total hours should increase under both higher government spending and lower taxes, we find no response (figure 6c). Interestingly, this lack of response results from both a positive response in total employment and a negative response in the number of hours worked per employee (not shown for brevity). For real wages, we find a gradual increase to a cumulative peak of around 3 percent compared to an otherwise similar economy that did not undergo any fiscal expansion. This suggests that the increase in labor demand dominates the increase in labor supply that may be associated with some of these policies (e.g., tax cuts on personal income).

The response of the inflation and interest rate is a little puzzling since none show an increase. Although this is a counterintuitive result, it should be noted that a negative relationship between inflation and fiscal shocks has been found in other studies (Canzoneri, Cumby, and Diba 2002; Canova and Pappa 2007; Fatas and Mihov 2001; Mountford and Uhlig 2009). Sebastian Ruth (2018) further argues that this response of policy rates to the fiscal stimulus does not reflect a direct reaction of monetary policy to the fiscal shock, but happens because fiscal shocks tend to occur in periods when there is a lack of inflationary pressure.

²⁴Not shown here for brevity is the response of net exports per se, which decline by up to 0.6 percent of GDP after two years. That is in line with Fred Bergsten and Joseph Gagnon (2017), who estimate the impact of the cyclically adjusted fiscal balance on the current account minus investment income (both relative to trend GDP) to be at above 0.5 for economies with high capital mobility (column 2 of table 2.1).

Figure 6: Impact of 1 percent of GDP fiscal expansion: GDP components, labor market, interest and inflation rates



3.5 Robustness Checks

If our series is truly exogenous there should be no need to control for other structural shocks. But fiscal expansions, like fiscal consolidations, often come as a package (Alesina, Favero, and Giavazzi 2019). The 1982 Reagan tax cuts, for instance, were accompanied by a deregulation push that, if anything, should also have had a positive impact on short-term growth. Similarly, the 2003 income tax cuts in Finland were implemented in the context of a broader agreement to moderate wages that might decrease domestic demand but help increase foreign demand. Our results could thus be affected by other structural shocks or accompanying policies, especially given the limited size of our sample. To help address this issue, we augment the local projection (LP) specification with changes in short-term interest rates to control for monetary policy (as in Romer and Romer (2010), Mertens and Ravn (2012), and Cloyne (2013)).²⁵ Doing this does not change results significantly (not shown for brevity).

More generally, the small size of our sample could make our results sensitive to the presence of outliers. To address this issue, we reestimate equation (9) but drop one episode at a time from our database. The thick red line in figure 7a shows our baseline estimate. The black dotted lines show the results one would obtain by dropping one episode at a time. Clearly, one episode seems to matter for the effect to converge to 1.5 in year 3 rather than in year 4, but none affects the shape of the IRF and its endpoint rests within a 0.5 percentage point window from the baseline specification.

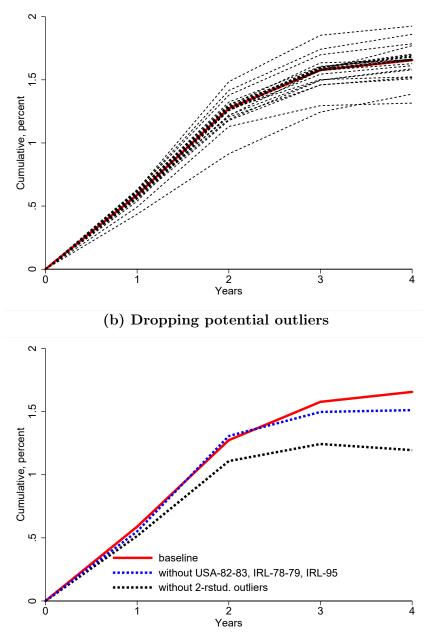
Potential outliers can also be spotted by examining the studentized residuals in our regressions. Thus, one should pay attention to studentized residuals that exceed +2 or -2. The black dotted line in figure 7b shows the results one would obtain by dropping 42 country-year observations that generate horizon 1 studentized residuals that exceed 2 in absolute value. However, out of the 42 such country-year observations only 3 represent observations with nonzero ΔSPB : the United States in 1982, Ireland in 1979, and Ireland in 1995, where the former two are parts of two-year fiscal expansions. Therefore, the blue dotted line in figure 7b further shows the results one would obtain by dropping the three fiscal expansions only, the United States in 1982–83, Ireland in 1978–79, and Ireland in 1995. In both tests, removing potential outliers leads to a downward shift in the IRF, but the response still exceeds 1 after two years.

One problem recently discussed in the literature is that narrative shocks, here ΔSPB_t^t , are measured with error because (i) different estimates of the size of the impulse from historical records (IMF, OECD) require judgment for choosing a single measure, (ii) small fiscal expansions are neglected and censored to zero, and (iii) estimates of the fiscal impulse are not always specified using the same metrics (e.g., as a percentage of GDP or potential

 $^{^{25}}$ We follow a similar timing convention to that used by these authors within a VAR framework.

Figure 7: Robustness checks

(a) Dropping one episode at a time



Note: 2-rstud. = studentized residuals exceeding +/-2. The lines indicate the cumulative percentage change in real GDP at years 1 to 4 relative to year 0 in response to a fiscal shock of 1 percentage point of GDP in year 1. Note for panel (a): The black dotted lines show the results one would obtain by dropping one fiscal episode at a time. Note for panel (b): The black dotted line shows the results one would obtain by dropping 42 country-year observations that generate studentized residuals that exceed 2 in absolute value. The blue dotted line shows the results one would obtain by dropping the three fiscal expansions: the US in 1982-83, Ireland in 1978-79 and Ireland in 1995.

 $GDP).^{26}$

Mertens and Ravn (2013) show that these measurement errors can create an attenuation bias, but that the narrative measure can be used as a proxy for the latent fiscal shock. For the narrative measure to be a good proxy, it needs to be (i) correlated with the structural fiscal shock and (ii) uncorrelated contemporaneously with all other structural shocks. Under these conditions, unbiased impulse responses can be obtained by estimating a proxy Structural VAR (Mertens and Ravn 2013), where a naïve measure of the change in fiscal policy (e.g., the change in tax revenues, government expenditures, or the deficit) is instrumented by the narrative measure. In our setup, this means using ΔSPB_t^t as a proxy for the latent fiscal shock and as an instrument for BFI_t^T . When using LP (see our equation (9)), the same proxy approach can be applied, providing that the narrative measure is uncorrelated with all other structural shocks at all leads and lags (Ramey 2016; Stock and Watson 2018). Under these assumptions, the causal effect of the fiscal expansions can be estimated via a standard two-stage least squares (2SLS) approach. The local projection-instrumental variable (LP-IV) methodology has the advantage of providing consistent estimates even in the presence of measurement error in explanatory variables, so long as the instrument (and its measurement error) is uncorrelated with any measurement error in the explanatory variables. In our setup, this translates into the condition that ΔSPB_t^t should be uncorrelated with measurement error in BFI_t^T .

We implement LP-IV with 2SLS. Specifically, we regress in a first stage BFI_t^T on the narrative measure ΔSPB_t^t , on country (c_i^1) and time (τ_t^1) fixed-effects, and on its own lag. In a second stage, we regress GDP growth on the fitted values of BFI_t^T and the same set of controls than in the first step. The two stages of the 2SLS estimator are presented for illustrative purposes. Although its name reflects the fact this estimator can be calculated in a two-step procedure, it is, in fact, calculated in one step in Stata. Thus, there is no need to further adjust the standard errors. Figure 8 compares the baseline IRF (dotted blue line) with that obtained when using the narrative measure as a proxy of the true underlying structural shock (solid green line). As expected, the new point estimates are higher at all horizons but the difference between the two lines is economically small and not statistically significant.

 $^{^{26}}$ In the context of equation (2) these measurement errors are of M_0 type as they are not systematically related to output movements.

Stage 1:
$$BFI_{i,t}^{T} = c_{i}^{1} + \tau_{t}^{1} + \varphi^{1}\Delta Y_{i,t-1} + \sum_{j=0}^{1} \delta_{j}^{1}\Delta SPB_{i,t-j}^{t-j} + \epsilon_{i,t}^{1}$$

$$BFI_{i,t-1}^{T} = c_{i}^{2} + \tau_{t}^{2} + \varphi^{2}\Delta Y_{i,t-1} + \sum_{j=0}^{1} \delta_{j}^{2}\Delta SPB_{i,t-j}^{t-j} + \epsilon_{i,t}^{2}$$
Stage 2: $Y_{i,t+h} - Y_{i,t-1} = \alpha_{i}^{h} + \lambda_{t}^{h} + \phi^{h}\Delta Y_{i,t-1} + \sum_{j=0}^{1} \beta_{j}^{h} \widehat{BFI_{i,t-j}^{T}} + u_{i,t+h}$

Another concern is that the multipliers we calculate are impact rather than cumulative multipliers. The impact multiplier is defined as the cumulative response of output up to a given year over the *initial* shift in fiscal policy. The cumulative multiplier compares the same cumulative response of output to the *cumulative* change in fiscal policy over the same period (Ramey 2016; Ramey 2019; Stock and Watson 2018). To obtain cumulative multipliers, one has only to replace BFI_t^T in stage 2 by its sum over the relevant horizon h. More specifically, we estimate the following equations:

Stage 1:
$$\sum_{k=0}^{h} BFI_{i,t+k}^{T} = c_{i}^{1h} + \tau_{t}^{1h} + \varphi^{1h} \Delta Y_{i,t-1} + \sum_{j=0}^{1} \delta_{j}^{1h} \Delta SPB_{i,t-j}^{t-j} + \epsilon_{i,t+h}^{1}$$
$$\sum_{k=0}^{h} BFI_{i,t-1+k}^{T} = c_{i}^{2h} + \tau_{t}^{2h} + \varphi^{2h} \Delta Y_{i,t-1} + \sum_{j=0}^{1} \delta_{j}^{2h} \Delta SPB_{i,t-j}^{t-j} + \epsilon_{i,t+h}^{2}$$
Stage 2:
$$\sum_{k=0}^{h} \Delta Y_{i,t+k} = \alpha_{i}^{h} + \lambda_{t}^{h} + \varphi^{h} \Delta Y_{i,t-1} + \sum_{j=0}^{1} \beta_{j}^{h} \overline{\sum_{k=0}^{h} BFI_{i,t-j+k}^{T}} + u_{i,t+h}$$

Figure 8 shows the difference between impact and cumulative (dashed purple line) multipliers. An important gap opens between the two types of multipliers in year 2, suggesting that the fiscal policy continues to be stimulative after the first year of the policy shift. As a result, the cumulative multiplier converges to around 1 after three years, a value one-third lower than that of the impact multiplier.

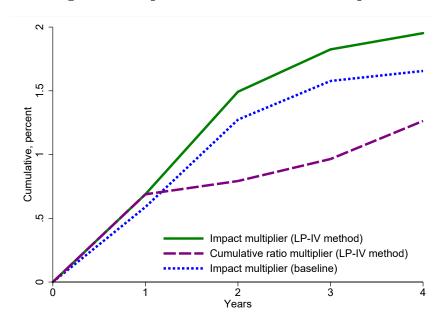


Figure 8: Impact and cumulative multipliers

Note: The lines indicate the cumulative percentage change in real GDP in years 1 to 4 relative to year 0 in response to a fiscal shock in year 1.

3.6 State Dependence: Fiscal Multipliers in Good and Bad Times

There are various reasons why the macroeconomic effects of fiscal policy may vary depending on the state of the economy. Most often cited is the view that fiscal policy may be more effective when the economy is in a slump and operating below capacity.²⁷ In this view, an increase in the budget deficit would not crowd out and might even crowd in private domestic spending when there is slack. This could arise because prices are less responsive in this environment (higher labor elasticity and lower markups, Hall (2009)) or because an economy with idle resources does not hit bottlenecks and capacity constraints (Gordon and Krenn 2010). It may also arise because credit constraints and financial frictions are countercyclical (Tagkalakis 2008; Canzoneri, Collard, Dellas, and Diba 2016) or because higher public sector employment leads to a milder increase in labor market tightness (Michaillat 2014). Finally, it could arise because of the lack of inflationary concerns and the muted response of monetary policy (Hall 2009).²⁸

The evidence is, however, mixed. Using a panel of OECD countries, Auerbach and Gorodnichenko (2013) and Steinar Holden and Victoria Sparrman (2018) find that higher government expenditures lead to a larger reduction in unemployment when the output gap

²⁷Also cited is whether the economy is moving from its peak to its trough (recession) or moving from its trough to its peak (expansion).

²⁸This is particularly the case when the economy is at the effective lower bound on interest rates (Christiano, Eichenbaum, and Rebelo 2011; Woodford 2011), but the point is more general.

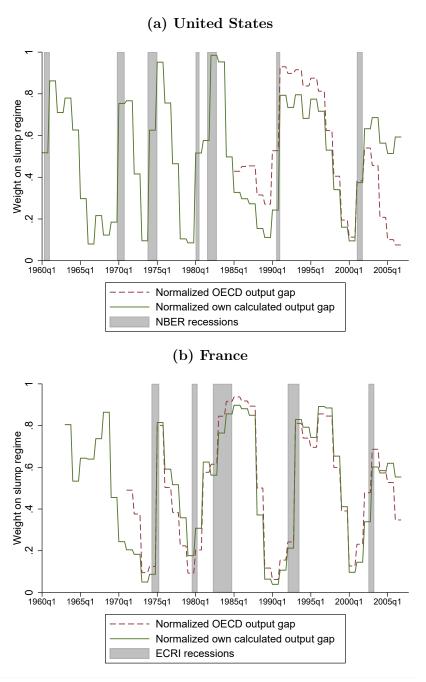
is negative. Steven Fazzari, James Morley, and Irina Panovska (2015) also find that the government spending multiplier is larger and more persistent whenever there is considerable economic slack. On the other hand, Michael Owyang, Valerie Ramey, and Sarah Zubairy (2013) and Ramey and Zubairy (2018) find no evidence of higher spending multipliers during periods of high unemployment in the United States and in subsequent research attribute the higher spending multiplier found for Canada to exceptional circumstances. For US taxes, Ruhollah Eskandari (2015) and Ufuk Demirel (2016) find that multipliers are actually smaller when unemployment is high than when it is low. Alesina, Favero, and Giavazzi (2019) do not find that state dependence matters for fiscal consolidations but do not consider define the state of the economy with a measure of slack. In contrast, Jordà and Taylor (2016) find strong state-dependence effects with a 1 percent of GDP fiscal consolidation associated with a reduction of real GDP by around 4 percent after five years when the economy is in a slump and no negative effect when it is in a boom.

3.6.1 Methodology

We contribute to this literature by investigating how the effects of fiscal stimulus vary depending on the level of slack in the economy. Methodologically, we follow Auerbach and Gorodnichenko (2012; 2013) and use a continuous measure of the state of the economy, expressed by the logistic transition function $\Lambda(z_t) = \frac{exp(-\rho z_t)}{1+exp(-\rho z_t)}$, which assigns to each state z at time t a value between 0 and 1 that can be interpreted as the probability of being in the bad state. An alternative approach would be to use the discrete threshold specification adopted by Owyang, Ramey, and Zubairy (2013), Ramey (2019), and Eskandari (2015), which replaces the transition function $\Lambda(.)$ with an indicator function that takes the value of 1 if state z expressed by the unemployment rate is above a certain threshold and zero otherwise. The advantage of the Auerbach and Gorodnichenko (2012; 2013) approach is that the statistical inference for each regime is effectively based on a larger set of observations than when using a discrete threshold specification and is thus particularly appropriate in our case.

We define the state variable z by estimating an output gap as the deviation of the logarithm of real GDP from a trend obtained using a Hodrick-Prescott (HP) filter with the smoothing parameter of 100, which is typical for annual data. Before entering the equation, the gap series is standardized and then transformed by the transition function, $\Lambda(.)$. The transition parameter ρ , which stretches the function around the 0.5-level along the Y-axis, is set to 1.5 as in Auerbach and Gorodnichenko (2013). Figure 9 illustrates the dynamics of $\Lambda(z_t)$ for the United States and France against that obtained with the OECD measure of the output gap. It also displays as shaded areas recession dates obtained from the National Bureau of Economic Research (NBER) for the US and from the Economic Cycle Research Institute (ECRI) for France. The first thing to note is that our simple HP filtering generates

Figure 9: Estimated weight on the slump regime: using own calculated or OECD output gap ${\bf P}$



a measure of economic slack that is very close to that obtained by using the output gap as calculated by the OECD when the two series overlap. In fact, the within correlation between the two measures is equal to 0.87. The second thing to note is that the probability of the bad state typically spikes following a recession. For recessions followed by strong recoveries, the measure goes down fairly quickly. For other recessions, the increase in slack is persistent. Appendix table B.5 gives the probability of a slump for each exogenous large fiscal expansion in our sample.

We estimate the following state-dependent local projections for the same variables as equation (9):

$$Y_{i,t+h} - Y_{i,t-1} = \alpha_i^h + \lambda_t^h + (1 - \Lambda(z_{i,t})) \left[\phi_B^h \Delta Y_{i,t-1} + \sum_{j=0}^1 \beta_{j,B}^h \Delta SPB_{i,t-j}^{t-j} \right]$$

$$+ \Lambda(z_{i,t}) \left[\phi_S^h \Delta Y_{i,t-1} + \sum_{j=0}^1 \beta_{j,S}^h \Delta SPB_{i,t-j}^{t-j} \right] + \delta^h \Lambda(z_{i,t}) + u_{i,t+h}$$
(10)

The dynamics are constructed by varying the horizon h of the dependent variable so that we can directly read the impulse responses from estimated $\{\hat{\beta}_{0,B}^h\}_{h=0}^H$ for booms (good state) and $\{\hat{\beta}_{0,S}^h\}_{h=0}^H$ for slumps (bad state).

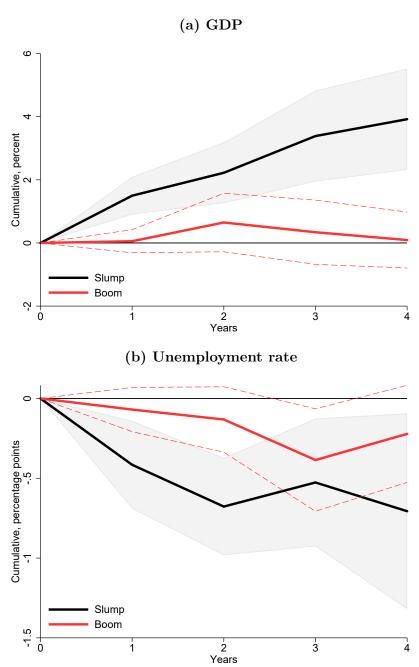
3.6.2 Results

We start by reporting in figure 10 the IRFs for (the natural logarithm of) real GDP and the unemployment rate for the slump and boom regimes. For all IRFs, the shock corresponds to fiscal impulse measure ΔSPB that happens in year 1 and the forecast horizon is 4 years.

According to our estimates, a 1 percent of GDP fiscal expansion is associated with an over 3 percent peak cumulative increase in GDP during a slump. The increase in output is statistically significant at the 5 percent level at each horizon after the fiscal shock. The size of the effect is consistent (in absolute value) with that obtained for fiscal consolidations by Jordà and Taylor (2016), who show that a 1 percent of GDP fiscal consolidation is associated with a reduction of real GDP by around 4 percent after five years when the economy is in a slump. Like them, we also find that the effects are not statistically different from zero when the economy is in a boom. In addition, the difference between our slump and boom IRFs is statistically significant at the 10 percent level at horizons 1 and 4. The difference in results between slump and boom is less clear for the unemployment rate. The point estimates in a boom are smaller (in absolute value) than those in a slump and are not statistically different from zero in the first two years of the shock.

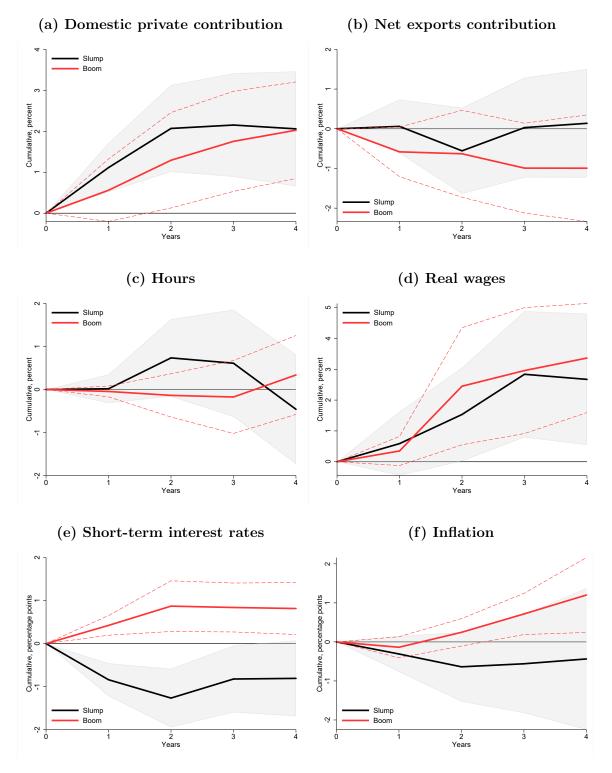
The top two panels of figure 11 show the contribution of private final demand and net exports to real GDP growth. They show that the difference in GDP between slump and boom is due not so much to a higher response of private consumption and private investment

Figure 10: Impact of 1 percent of GDP fiscal expansion: GDP and unemployment rate



Note for panel (a): The solid lines indicate the cumulative percentage changes in real GDP in years 1 to 4 relative to year 0 in response to a fiscal shock of 1 percentage point of GDP in year 1. Note for panel (b): The solid lines indicate the cumulative percentage point changes in the unemployment rate. Note for panels (a) and (b): The dashed lines and shaded areas represent one standard–error bands.

Figure 11: Impact of 1 percent of GDP fiscal expansion: GDP contributions, labor markets, interest and inflation rates



during a slump as to the stronger offset from net exports during a boom. Another important difference revealed by the state-dependent results is the behavior of inflation and interest rates. Fiscal expansion in a slump does not appear to generate inflationary concerns. Rather than counteracting the fiscal expansions, monetary policy accommodates or even supports it by also providing stimulus in the form of lower rates. The opposite happens during a boom, thereby dampening the impact of fiscal stimulus.

A first robustness check consists of modifying the timing convention to define the state of the economy. Mario Alloza 2017 and Ramey and Zubairy (2018), for instance, show that using only past data to define the state of the economy reverses the findings of high multipliers in recessions and low multipliers in expansions of Auerbach and Gorodnichenko (2012). To check for this possibility, we restimate equation (10) but use the output gap in t-1 (i.e., one year prior to the fiscal stimulus) rather than in t to define the state of the economy. Appendix figure C.1 shows the IRFs for GDP and the interest rate. Modifying the timing convention does lower the point estimates for slumps and brings them closer to those found for booms. But, and in contrast to the effects for booms, the effects for slumps remain statistically different from zero at the 10 percent level during the first three years.

A second robustness check consists of changing the way we constructed the output gap. In particular, we can check whether our results depend on the specific value of the smoothing parameter chosen to extract an HP trend. Auerbach and Gorodnichenko (2013) and Eskandari (2015), for example, use larger values of the smoothing parameter to get a trend that is only slowly moving. In their sample, it is necessary to avoid treating the Great Recession as a reduction in potential output rather than as a reduction in cyclical output. Reflecting the fact that our sample ends before the Great Recession starts, the particular value of the smoothing parameter chosen is not crucial with our sample, values between 6.25 or 100, which are typical for annual frequency (Ravn and Uhlig 2002), and 5,000 give essentially the same results.

A third robustness check consists of changing the parameter ρ of the transition function that is set to 1.5 as in Auerbach and Gorodnichenko (2013) in our baseline specification. With a smoothing parameter of 100 for the HP filter, we doubled the value of ρ to 3 and found that it did not substantially affect our state-dependent results for real GDP.

3.6.3 Multipliers in Deep Slumps and Strong Booms

One recently considered hypothesis is that fiscal multipliers may be particularly large in severe economic conditions (Caggiano, Castelnuovo, Colombo, and Nodari 2015). According to this view, it is not so much whether the economy is above or below potential that matters, but by how much.²⁹ To test the hypothesis that fiscal multipliers behave differently in deep

 $^{^{29}\}mathrm{Caggiano},$ Castelnuovo, Colombo, and Nodari (2015) look at deep recessions and strong expansions.

slumps and in strong booms, we switch to a discrete threshold specification of equation (10). Specifically, we begin by reestimating the equation separately for positive and negative (standardized) output gap z values as in Jordà and Taylor (2016). This would be almost equivalent to replacing the transition function $\Lambda(.)$ with an indicator function (dummy variable) that takes the value of 1 if the output gap is negative and 0 otherwise.³⁰ Doing this allows us not only to compare results based on two different econometric specifications, but also to then modify the value of the threshold considered.

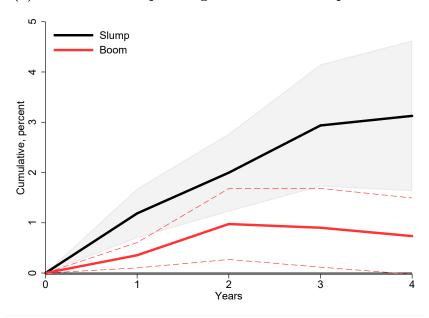
The top panel of figure 12 compares the IRFs for booms and slumps as was done in the top panel of figure 10, but using a specification where a threshold rather than a continuous variable is used to define the state. The difference in results between slump and boom becomes less stark, although a 1 percent of GDP fiscal expansion conducted in a slump is still associated with a sizable cumulative increase in real GDP of over 2.5 percent after three years.

We next reestimate the discrete-state specification for stronger booms and deeper slumps. A period where the value of the standardized output gap z is above 0.5 or 0.75 (i.e., 0.5 or 0.75 standard deviation above its mean) is defined as a strong boom. Similarly, a period where the value of the standardized output gap z is below -0.5 or -0.75 (i.e., 0.5 or 0.75 standard deviation below its mean) is defined as a deep slump. When moving from 0 to 0.75 standardized output gap, the number of boom country-year observations declines from 24 to 16, and the number of slump observations declines from 11 to 7. The bottom panel of figure 12 shows the difference between IRFs in booms and slumps for different values of the discrete output gap threshold. Thus, the difference between IRFs in the two states gets bigger for strong booms and deep slumps. While the difference is not statistically significant at conventional levels for the 0 threshold, it becomes statistically significant at the 5 percent level over most horizons for the 0.75 threshold.

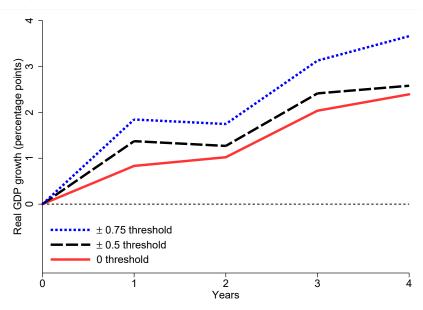
 $^{^{30}}$ This is not fully equivalent because equation (10) presumes homogeneous time effects and homogeneous error variances across the two discrete states.

Figure 12: Impact of 1 percent of GDP fiscal expansion on GDP

(a) Booms vs slumps using a discrete-state specification



(b) Difference between strong booms and deep slumps



Note for panel (a): The solid lines indicate the cumulative percentage changes in real GDP in years 1 to 4 relative to year 0 in response to a fiscal shock of 1 percentage point of GDP in year 1. Boom state takes into account episodes that happen when the standardized output gap is above 0, and slump means it is below 0. The dashed lines and shaded area represent one standard—error bands. Note for panel (b): The lines show the difference between impulse response functions in booms and slumps for different values of the threshold. The threshold is defined in terms of the standardized output gap. When moving it from 0 to 0.75 in absolute value, the number of boom country-year observations declines from 24 to 16, and the number of slump observations declines from 11 to 7.

4 Conclusion

In this paper, we construct a new dataset of large fiscal expansions. To do this, we introduce a hybrid methodology that complements the standard statistical approach to identifying episodes of large fiscal changes with narrative evidence. We find evidence that multipliers are in general above 1 and state-dependent. We also show that the problems of measurement errors and reverse causality that typically complicate efforts to determine the effects of changes in fiscal policy are, if anything, likely to be bigger for stimulus than for austerity.

Several improvements and extensions can be envisaged. First, the dataset could be extended by successively decreasing the threshold beyond which a decline in the cyclically adjusted primary balance is considered large. Eventually, this would make the dataset exhaustive. Second, the analysis could be refined by documenting the exact composition of each episode between tax cuts and expenditure increases and investigating whether this distinction is as important for fiscal expansions as for fiscal consolidations (Alesina, Favero, and Giavazzi 2019).

References

- Abiad, Abdul, Davide Furceri, and Petia Topalova. 2016. The Macroeconomic Effects of Public Investment: Evidence from Advanced Economies. *Journal of Macroeconomics* vol. 50, no. C, pp. 224–240.
- Alesina, Alberto. 2010. My answer to The Economist. URL: https://scholar.harvard.edu/alesina/content/my-answer-economist (visited on 07/03/2019).
- Alesina, Alberto and Silvia Ardagna. 2010. Large Changes in Fiscal Policy: Taxes versus Spending. *Tax Policy and the Economy, Volume 24*. Ed. by Jeffrey R. Brown. Cambridge, MA: National Bureau of Economic Research.
- Alesina, Alberto, Omar Barbiero, Carlo Favero, Francesco Giavazzi, and Matteo Paradisi. 2017. The Effects of Fiscal Consolidations: Theory and Evidence. NBER Working Paper 23385. Cambridge, MA: National Bureau of Economic Research.
- Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2015. The Output Effect of Fiscal Consolidations. *Journal of International Economics* vol. 96, pp. S19–S42.
- Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019. Austerity: When It Works and When It Doesn't. Princeton University Press.
- Alesina, Alberto and Roberto Perotti. 1995. Fiscal Expansions and Fiscal Adjustments in OECD Countries. NBER Working Paper 5214. Cambridge, MA: National Bureau of Economic Research.
- Alloza, Mario. 2017. Is fiscal policy more effective in uncertain times or during recessions? Working Paper 1730. Madrid: Banco de España.
- Arellano, Manuel and Stephen Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* vol. 58, no. 2, pp. 277–297.
- Auerbach, Alan and Yuriy Gorodnichenko. 2012. Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy* vol. 4, no. 2, pp. 1–27.
- Auerbach, Alan and Yuriy Gorodnichenko. 2013. Fiscal Multipliers in Recession and Expansion. Fiscal Policy after the Financial Crisis. Ed. by Alberto Alesina and Francesco Giavazzi. University of Chicago Press.
- Barnichon, Regis and Christian Matthes. 2016. Understanding the Size of the Government Spending Multiplier: It's in the Sign. CEPR Discussion Papers 11373. Washingon, DC: Center for Economic and Policy Research.
- Beetsma, Roel, Oana Furtuna, and Massimo Giuliodori. 2017. Revenue- versus Spending-Based Consolidation Plans: The Role of Follow-Up. CEPR Discussion Paper 12133. Washington, DC: Center for Economic and Policy Research.

- Bergsten, Fred and Joseph Gagnon. 2017. Currency Conflict and Trade Policy: A New Strategy for the United States. Columbia University Press.
- Blagrave, Patrick, Giang Ho, Ksenia Koloskova, and Esteban Vesperoni. 2017. Fiscal Spillovers: The Importance of Macroeconomic and Policy Conditions in Transmission. IMF Spillover Note 17/02. Washington, DC: International Monetary Fund.
- Blanchard, Olivier. 1990. Suggestions for a New Set of Fiscal Indicators. OECD Economics Department Working Paper 79. Paris: OECD Publishing.
- Blanchard, Olivier J, Alvaro Leandro, Silvia Merler, and Jeromin Zettelmeyer. 2018. Impact of Italy's Draft Budget on Growth and Fiscal Solvency. Policy Briefs PB18-24. Washington, DC: Peterson Institute for International Economics.
- Blanchard, Olivier and Roberto Perotti. 2002. An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics* vol. 117, no. 4, pp. 1329–68.
- Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari. 2015. Estimating Fiscal Multipliers: News from a Non-linear World. *The Economic Journal* vol. 125, no. 584, pp. 746–76.
- Canova, Fabio and Evi Pappa. 2007. Price Differentials in Monetary Unions: The Role of Fiscal Shocks. *The Economic Journal* vol. 117, no. 520, pp. 713–37.
- Canzoneri, Matthew, Fabrice Collard, Harris Dellas, and Behzad Diba. 2016. Fiscal Multipliers in Recessions. *The Economic Journal* vol. 126, no. 590, pp. 75–108.
- Canzoneri, Matthew, Robert Cumby, and Behzad Diba. 2002. Should the European Central Bank and the Federal Reserve be concerned about fiscal policy? *Proceedings Economic Policy Symposium Jackson Hole*, pp. 333–89.
- Carnot, Nicolas and Francisco de Castro. 2015. The Discretionary Fiscal Effort: An Assessment of Fiscal Policy and Its Output Effect. *Hacienda Pública Española* vol. 215, no. 4, pp. 63–94.
- Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. 2011. When Is the Government Spending Multiplier Large? *Journal of Political Economy* vol. 119, no. 1, pp. 78–121.
- Cloyne, James. 2013. Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom. *American Economic Review* vol. 103, no. 4, pp. 1507–28.
- Cohen-Setton, Jérémie and Natacha Valla. 2010. Measuring Fiscal Adjustments: Behind the Scenes. Goldman Sachs European Weekly Analyst, no. 10/22.
- Darvas, Zsvolt. 2013. Mind the Gap! And the Way Structural Budget Balances Are Calculated. *Bruegel Blog*.

- Demirel, Ufuk. 2016. The Short-Term Effects of Tax Changes: Evidence for State Dependence. Working Paper 51824. Washington, DC: Congressional Budget Office.
- Driscoll, John and Aart Kraay. 1998. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics* vol. 80, no. 4, pp. 549–60.
- Eskandari, Ruhollah. 2015. Three Essays in Tax Policy, Macroeconomics, and Corporate Policies. Chapter 1. State-Dependent Macroeconomic Effects of Tax Changes: Estimates Based on Narrative Records in the U.S., Lombardy Advanced School of Economic Research. PhD thesis.
- Fatas, Antonio and Ilian Mihov. 2001. Fiscal Policy and Business Cycles: An Empirical Investigation. *Moneda y Crédito* vol. 212, pp. 167–210.
- Fazzari, Steven, James Morley, and Irina Panovska. 2015. State-dependent effects of fiscal policy. Studies in Nonlinear Dynamics & Econometrics vol. 19, no. 3, pp. 285–315.
- Frees, Edward. 1995. Assessing cross-sectional correlation in panel data. *Journal of Econometrics* vol. 69, no. 2, pp. 393–414.
- Furceri, Davide and Bin Grace Li. 2017. The Macroeconomic (and Distributional) Effects of Public Investment in Developing Economies. IMF Working Paper 17/217. Washington, DC: International Monetary Fund.
- Girouard, Nathalie and Christophe André. 2005. Measuring Cyclically-Adjusted Budget Balances for OECD Countries. OECD Economics Department Working Paper 434. Paris: OECD Publishing.
- Gordon, Robert J. and Robert Krenn. 2010. The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers. NBER Working Paper 16380. Cambridge, MA: National Bureau of Economic Research.
- Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. Expansionary Austerity? International Evidence. *Journal of the European Economic Association* vol. 12, no. 4, pp. 949–68.
- Hall, Robert. 2009. By How Much Does GDP Rise If the Government Buys More Output? *Brookings Papers on Economic Activity* vol. 40, no. 2 (Fall), pp. 183–249.
- Holden, Steinar and Victoria Sparrman. 2018. Do Government Purchases Affect Unemployment? *The Scandinavian Journal of Economics* vol. 120, no. 1, pp. 124–58.
- IMF. 2010. Will It Hurt? Macroeconomic Effects of Fiscal Consolidation. URL: https://www.imf.org/external/pubs/ft/weo/2010/02/ (visited on 07/03/2019).

- IMF. 2013. Reassessing the Role and Modalities of Fiscal Policy in Advanced Economies. IMF Policy Paper. Washington, DC: International Monetary Fund.
- Jordà, Oscar. 2005. Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review* vol. 95, no. 1, pp. 161–182.
- Jordà, Òscar and Alan Taylor. 2016. The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy. *The Economic Journal* vol. 126, no. 590, pp. 219–55.
- Journard, Isabelle, Makoto Minegishi, Christophe André, Chantal Nicq, and Robert Price. 2008. Accounting for One-off Operations When Assessing Underlying Fiscal Positions. OECD Economics Department Working Paper 642. Paris: OECD Publishing.
- Koske, Isabell and Nigel Pain. 2008. The Usefulness of Output Gaps for Policy Analysis. OECD Economics Department Working Paper 621. Paris: OECD Publishing.
- Mertens, Karel and Morten Ravn. 2012. Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks. *American Economic Journal: Economic Policy* vol. 4, no. 2, pp. 145–81.
- Mertens, Karel and Morten Ravn. 2013. The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review* vol. 103, no. 4, pp. 1212–47.
- Mertens, Karel and Morten Ravn. 2014. A reconciliation of SVAR and narrative estimates of tax multipliers. *Journal of Monetary Economics* vol. 68, no. S, pp. S1–S19.
- Michaillat, Pascal. 2014. A Theory of Countercyclical Government Multiplier. American Economic Journal: Macroeconomics vol. 6, no. 1, pp. 190–217.
- Mountford, Andrew and Harald Uhlig. 2009. What are the effects of fiscal policy shocks? *Journal of Applied Econometrics* vol. 24, no. 6, pp. 960–92.
- Nickell, Stephen. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica* vol. 49, no. 6, pp. 1417–26.
- OECD. 2006. Should Measures of Fiscal Stance Be Adjusted for Terms of Trade Effects? OECD Economics Department Working Paper 519. Paris: OECD Publishing.
- Owyang, Michael, Valerie Ramey, and Sarah Zubairy. 2013. Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data. *American Economic Review* vol. 103, no. 3, pp. 129–34.

- Pesaran, M. Hashem. 2004. General Diagnostic Tests for Cross Section Dependence in Panels. Cambridge Working Papers in Economics. Faculty of Economics, University of Cambridge.
- Price, Robert and Thai-Thanh Dang. 2011. Adjusting Fiscal Balances for Asset Price Cycles. OECD Economics Department Working Paper 868. Paris: OECD Publishing.
- Princen, Savina, Gilles Mourre, Dario Paternoster, and George-Marian Isbasoiu. 2013. Discretionary Tax Measures: Pattern and Impact on Tax Elasticities. European Economy-Economic Papers 2008-2015 499. Directorate General Economic and Financial Affairs (DG ECFIN). Brussels: European Commission.
- Ramey, Valerie. 2016. Macroeconomic Shocks and Their Propagation. Vol. 2. Handbook of Macroeconomics. Elsevier. Chap. 0, pp. 71–162.
- Ramey, Valerie. 2019. Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research? *Journal of Economic Perspectives* vol. 33, no. 2, pp. 89–114.
- Ramey, Valerie and Sarah Zubairy. 2018. Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. *Journal of Political Economy*. eprint: https://doi.org/10.1086/696277. (Visited on 07/02/2019).
- Ravn, Morten and Harald Uhlig. 2002. On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics* vol. 84, no. 2, pp. 371–75.
- Riera-Crichton, Daniel, Carlos A. Vegh, and Guillermo Vuletin. 2016. Tax Multipliers: Pitfalls in Measurement and Identification. *Journal of Monetary Economics* vol. 79, pp. 30–48.
- Romer, Christina and David Romer. 2010. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review* vol. 100, no. 3, pp. 763–801.
- Ruth, Sebastian. 2018. Fiscal stimulus and systematic monetary policy: Postwar evidence for the United States. *Economics Letters* vol. 173, no. C, pp. 92–96.
- Stock, James and Mark Watson. 2018. Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal* vol. 128, no. 610, pp. 917–48.
- Tagkalakis, Athanasios. 2008. The effects of fiscal policy on consumption in recessions and expansions. *Journal of Public Economics* vol. 92, no. 5-6, pp. 1486–1508.
- Woodford, Michael. 2011. Simple Analytics of the Government Expenditure Multiplier. American Economic Journal: Macroeconomics vol. 3, no. 1, pp. 1–35.

Yang, Weonho, Jan Fidrmuc, and Sugata Ghosh. 2015. Macroeconomic Effects of Fiscal Adjustment: A Tale of Two Approaches. *Journal of International Money and Finance* vol. 57, pp. 31–60.

A Implementation of the Blanchard Fiscal Impulse

We mostly use data on general government accounts from the OECD *Economic Outlook* 100 of June 2016. For some countries, this release does not go back far enough. Thus, for Australia, Austria, France, and Ireland, we use *Economic Outlook* N. 84 of June 2008. For Germany, we use *Economic Outlook* N. 76 of June 2004. As we choose to identify episodes up to 2008 it does not shorten our sample much.

Our calculation of the Blanchard fiscal impulse (BFI) follows closely the implementation approach of Alesina and Perotti (1995) and Alesina and Ardagna (2010). In particular, we first isolate the components of government revenues and expenditures that need to be adjusted for cyclical variations. Once these components are adjusted according to Blanchard (1990)'s idea, cyclically adjusted revenues and expenditures net of capital outlays are obtained. From this, we obtain a measure of the cyclically adjusted primary balance that we call BFI_t^T .

The following variables are used:

- CAPOG: Net capital outlays of the government, value
- CGAA: Government final consumption expenditure, value, appropriation account
- GGINTP: Gross government interest payments, value
- GGINTR: Gross government interest receipts, value
- SSPG: Social security benefits paid by general government, value
- SSRG: Social security contribution received by general government, value
- TIND: Taxes on production and imports, value
- TOCR: Other current receipts, general government, value
- TY: Total direct taxes, value
- YPEPG: Property income paid by government, value
- YPERG: Property income received by government, value
- YPG: Current disbursements, general government, value (CGAA + SSPG + YPEPG + YPOTG)
- YPOTG: Other current outlays, general government, value
- YRG: Current receipts, general government value (TIND + TY + YPERG + SSRG + TOCR)

Step 1: Calculate expenditures and revenues that need to be cyclically adjusted. For each country, define R^1 , the portion of government revenues that needs to be cyclically adjusted, as $R_t^1 = TY_t + TIND_t + SSRG_t$ (the sum of total direct taxes, taxes on production and imports and received social security contributions). For government expenditures, only paid social security benefits (SSPG) are included as expenditures that need to be adjusted for cyclical variations.

Step 2: Cyclically adjust the revenue and expenditure sides. For each country we regress R^1 and SSPG on the previous year's unemployment rate and on a time trend. From the regression we calculate the adjusted revenue and adjusted expenditure variables R_U and $SSPG_U$.

Step 3: Construct adjusted total revenue and adjusted total expenditure. Adjusted total revenue (R_t^*) is obtained by adding property income received by the government net of interest rate receipts and other receipts to $R_{U,t}$. Adjusted total expenditure (G_t^*) is obtained by adding final consumption of expenditure, property income paid by the government net of interest payments, net capital outlays and other outlays to $SSPG_{U,t}$.

- $R_t^* = R_{U,t} + (YPERG_t GGINTR_t) + TOCR_t$
- $G_t^* = SSPG_{U,t} + CGAA_t + (YPEPG_t GGINTP_t) + YPOTG_t + CAPOG_t$

Step 4: Calculate BFI_t^T as a measure of the change of the cyclically adjusted primary balance.

$$BFI_t^T = [R_t^* - G_t^*] - [R_{t-1} - G_{t-1}]$$

where the (unadjusted) total revenue and total expenditure of the government are defined as follow:

- $R_t = TY_t + TIND_t + SSRG_t + (YPERG_t GGINTR_t) + TOCR_t$
- $G_t = SSPG_t + CGAA_t + (YPEPG_t GGINTP_t) + YPOTG_t + CAPOG_t$

When our calculated BFI measure is compared to the change in the CAPB as calculated by the OECD, the variable NLGXA from the OECD $Economic\ Outlook$ is used.

Special cases: Editions N. 84 and N. 76 of the OECD *Economic Outlook*, which we use to obtain data for Australia, Austria, France, Germany, and Ireland do not contain data on interest payments/receipts (GGINTP/ GGINTR). For these countries, interest payments are thus not removed from adjusted revenues and expenditures. This data limitation is, however, unlikely to affect our findings because of our focus on particularly large shifts in the BFI measure. Data on social security contributions received by the government are also missing for Australia. In this case, adjustable revenue is thus $R_t = TY_t + TIND_t$.

B Tables

Table B.1: Mean and standard deviation of changes in cyclically adjusted fiscal measures (% of GDP)

Country	BFI measured as in Alesina and Perotti (1995)				$\Delta CAPB_t^T$ measured by OECD					
	# of obs.	μ_i	σ_i	1-yr thld	2-yr thld	# of obs.	μ_i	σ_i	1-yr thld	2-yr thld
Australia	42	.04	1.08	-1.05	-1.59	35	.07	.93	85	-1.32
Austria	46	.02	1.28	-1.25	-1.89	36	07	1.22	-1.29	-1.90
Belgium	36	.15	1.81	-1.66	-2.56	21	.11	1.22	-1.11	-1.72
Canada	36	.07	1.32	-1.24	-1.90	34	0	1.25	-1.26	-1.88
Denmark	35	.16	1.59	-1.43	-2.23	21	.06	1.22	-1.17	-1.78
Finland	46	.10	1.79	-1.69	-2.59	21	.01	1.65	-1.64	-2.47
France	43	03	.88	90	-1.34	34	08	.82	89	-1.30
Germany	46	02	1.36	-1.38	-2.06	35	02	1.09	-1.10	-1.64
Ireland	46	.11	1.66	-1.55	-2.38	26	.59	1.53	94	-1.70
Italy	46	03	1.38	-1.41	-2.10	21	.16	1.18	-1.02	-1.61
Japan	46	.12	1.32	-1.21	-1.87	21	11	1.73	-1.84	-2.70
Netherlands	37	.06	1.84	-1.77	-2.69	21	02	2.26	-2.27	-3.40
Portugal	29	02	1.62	-1.65	-2.46	21	18	1.43	-1.61	-2.32
Spain	29	.22	1.10	88	-1.43	21	.26	.98	71	-1.20
Sweden	46	.07	2.03	-1.96	-2.98	21	.13	2.19	-2.06	-3.16
United Kingdom	36	16	1.34	-1.49	-2.16	26	05	1.16	-1.21	-1.79
United States	46	03	.94	97	-1.44	21	01	1.03	-1.04	-1.55

Note: BFI = Blanchard fiscal impulse, $\Delta CAPB_t^T = \text{Change in cyclically adjusted primary balance}$, $\mu_i = \text{the mean of country } i$, $\sigma_i = \text{the standard deviation of country } i$, thid = threshold.

Table B.2: List of large declines in cyclically adjusted fiscal measures through 2006

Country (start of sample)	Year
Australia (1965)	1965, 1966, 1974, 1975, 1982, 1983, 1984*, 1985*,
	1989, 1990, 1991, 1992, 2000, 2001, 2004*
Austria (1961)	1963,1967,1975,1976,1993,2004
Belgium (1971)	$1972,\ 1975,\ 1981,\ 1983,\ 1988^*,\ 2002,\ 2003,\ 2005$
Canada (1971)	$1975, 1977^*, 1978^*, 1982, 1983, 1991, 2001, 2002$
Denmark (1972)	1974, 1975, 1989*, 1990*, 1991*, 1994*, 2006*
Finland (1961)	1963, 1978, 1979, 1982, 1983, 1987, 1990, 1991, 1992, 2001, 2002, 2003
France (1964)	1975, 1978, 1981, 1988*, 1990*, 1992, 1993, 2002, 2003
Germany (1961)	1974, 1975, 1987, 1988*, 1990, 1991, 1995*, 2000*, 2001
Ireland (1961)	1974, 1975, 1978, 1979, 1980, 1981, 1982, 1985*, 1990*, 1995, 2001, 2002
Italy (1961)	1965, 1971, 1972, 1975, 1981, 1984, 1985, 1998*, 2000*, 2001
Japan (1961)	1962, 1975, 1992, 1993, 1998
Netherlands (1970)	1975, 1995, 2001, 2002, 2003
Portugal (1978)	1978, 1987, 1990, 1991, 1993, 1998, 2004
Spain (1978)	1978, 1980, 1981, 1982, 1989*, 1990, 1991, 1993, 2000*, 2001*
Sweden (1961)	1977, 1978, 1979, 1980, 1991, 1992, 2001, 2002
United Kingdom (1971)	1971, 1972, 1973, 1983*, 1990, 1992, 1993, 2001, 2002, 2003
United States (1961)	1966, 1967, 1970, 1975, 1982, 1983, 1992*, 2001, 2002, 2003

Note: Starred episodes (*) were not identified using the methodology described in the text, but with the cyclically adjusted primary balance $(\Delta CAPB_t^T)$ as calculated by the OECD.

Table B.3: List of large declines in cyclically adjusted fiscal measures due to policy changes or other reasons (struck through)

Country (start of sample)	Year		
Australia (1965)	1965 , 1966, 1974, 1975, 1982, 1983, 1984*, 1985*,		
	1989 , 1990 , 1991 , 1992 , 2000, 2001, 2004*		
Austria (1961)	$1963, 1967, 1975, 1976, \frac{1993}{2004}, \frac{2004}{2000}$		
Belgium (1971)	$1972,1975,\underline{1981},\underline{1983},\underline{1988^*},\underline{2002},\underline{2003},\underline{2005}$		
Canada (1971)	$1975, 1977^*, 1978^*, \underline{1982}, 1983, \underline{1991}, 2001, 2002$		
Denmark (1972)	$\underline{1974}, 1975, \underline{1989^*}, \underline{1990^*}, \underline{1991^*}, 1994^*, \underline{2006^*}$		
Finland (1961)	1963, 1978, 1979, 1982, 1983, 1987, 1990, 1991, 1992 , 2001, 2002, 2003		
France (1964)	1975, 1978 , 1981, 1988* , 1990*, 1992 , 1993 , 2002, 2003		
Germany (1961)	1974, 1975, 1987, 1988*, 1990, 1991, 1995* , 2000* , 2001		
Ireland (1961)	1974, 1975, 1978, 1979, 1980, 1981, 1982 , 1985*, 1990* , 1995, 2001, 2002		
Italy (1961)	1965, 1971, 1972, 1975, 1981 , 1984, 1985, 1998*, 2000*, 2001		
Japan (1961)	1962, 1975, 1992, 1993, 1998		
Netherlands (1970)	1975, 1995 , 2001, 2002 , 2003		
Portugal (1978)	1978 , 1987, 1990, 1991, 1993 , 1998 , 2004		
Spain (1978)	1978, 1980, 1981, 1982, 1989* , 1990 . 1991 , 1993 , 2000* , 2001*		
Sweden (1961)	$1977,1978,1979,\underline{1980},\underline{1991},\underline{1992},2001,2002$		
United Kingdom (1971)	1971, 1972, 1973, 1983*, 1990 , 1992, 1993, 2001, 2002, 2003		
United States (1961)	1966 , 1967, 1970, 1975, 1982, 1983, 1992*, 2001, 2002, 2003		

Note: Date means that the decline in a cyclically adjusted fiscal measure is not mostly due to a change in fiscal policy. Starred episodes (*) were not identified using the methodology described in the text, but with $\Delta CAPB_t^T$ as calculated by the OECD.

Table B.4: List of large declines in cyclically-adjusted fiscal measures due to endogenous motivations, exogenous motivations (bold), or other reasons (struck through)

Country (start of sample)	Year
Australia (1965)	1965 , 1966, 1974, 1975, 1982, 1983, 1984*, 1985*,
	1989, 1990, 1991, 1992, 2000 , 2001, 2004*
Austria (1961)	1963 , 1967, 1975, 1976 , 1993 , 2004
Belgium (1971)	$1972,\ 1975,\ 1981,\ 1983,\ 1988^{*},\ 2002,\ 2003,\ 2005$
Canada (1971)	$1975,1977^*,1978^*,\underline{1982},1983,\underline{1991},2001,2002$
Denmark (1972)	$\underline{1974}, 1975, \underline{1989^*}, \underline{1990^*}, \underline{1991^*}, 1994^*, \underline{2006^*}$
Finland (1961)	1963, 1978, 1979, 1982, 1983, 1987, 1990, 1991, 1992 , 2001, 2002, 2003
France (1964)	1975, 1978 , 1981, 1988* , 1990* , 1992 , 1993 , 2002 , 2003
Germany (1961)	1974, 1975, 1987, 1988*, 1990 , 1991 , 1995* , 2000* , 2001
Ireland (1961)	1974, 1975, 1978 , 1979 , 1980, 1981, 1982 , 1985*, 1990* , 1995 , 2001 , 2002
Italy (1961)	1965, 1971, 1972, 1975, 1981 , 1984 , 1985 , 1998* , 2000* , 2001
Japan (1961)	1962, 1975, 1992, 1993, 1998
Netherlands (1970)	1975, 1995 , 2001 , 2002 , 2003
Portugal (1978)	1978, 1987 , 1990 , 1991 , 1993 , 1998 , 2004
Spain (1978)	1978 , 1980, 1981, 1982, 1989* , 1990 , 1991 , 1993 , 2000* , 2001*
Sweden (1961)	$1977,1978,1979,\underline{1980},\underline{1991},\underline{1992},2001,2002$
United Kingdom (1971)	1971, 1972, 1973, 1983* , 1990 , 1992, 1993, 2001 , 2002 , 2003
United States (1961)	1966 , 1967, 1970, 1975, 1982 , 1983 , 1992*, 2001, 2002 , 2003

Note: Date means that the decline in a cyclically adjusted fiscal measure is not mostly due to a change in fiscal policy. date means that it is due to a policy change motivated by endogenous reasons (i.e., countercyclical reasons). date means that it is due to a change in fiscal policy motivated by exogenous reasons. Starred episodes (*) were not identified using the methodology described in the text, but with $\Delta CAPB_t^T$ as calculated by the OECD.

Table B.5: Probability of an episode occurring in a bad state

Country	Year	$\Lambda(z_t)$	$\Lambda(z_{t-1})$
Australia	2000	.21	.19
Australia	2004	.48	.43
Austria	1963	.72	.64
Austria	1976	.39	.71
Canada	2001	.23	.09
Finland	2003	.55	.37
France	1990	.04	.06
France	2002	.34	.14
Germany	1990	.20	.69
Germany	1991	.02	.20
Germany	2001	.16	.23
Ireland	1978	.08	.33
Ireland	1979	.12	.08
Ireland	1995	.87	.97
Ireland	2001	.09	.04
Ireland	2002	.13	.09
Italy	1984	.86	.92
Italy	1985	.82	.86
Italy	1998	.64	.61
Italy	2000	.17	.66
Italy	2001	.16	.17
Netherlands	2001	.05	.03
Portugal	1987	.79	.94
Portugal	1990	.11	.13
Portugal	1991	.07	.11
Spain	1978	.24	.16
Sweden	1977	.84	.26
Sweden	1978	.85	.84
Sweden	2001	.27	.11
Sweden	2002	.46	.27
United Kingdom	1983	.82	.94
United Kingdom	2001	.38	.35
United Kingdom	2002	.49	.38
United Kingdom	2003	.40	.49
United States	1982	.98	.58
United States	1983	.95	.98
United States	2002	.63	.37
United States	2003	.69	.63

Note: A value of ρ needs to be chosen to evaluate the $\Lambda(z_t) = \frac{exp(-\rho z_t)}{1 + exp(-\rho z_t)}$ function, which estimates the probability of a bad state. Following Auerbach and Gorodnichenko (2013) we use $\rho = 1.5$.

C Figure

Figure C.1: State-dependent multipliers and monetary policy response: using t-1 gap to define the state.

