PSL Quarterly Review

vol. 73 n. 294 (September 2020)

Article

Inside the IMF "mea culpa": A panel analysis on growth forecast errors and Keynesian multipliers in Europe

EMILIANO BRANCACCIO and FABIANA DE CRISTOFARO*

Abstract:

The tests carried out by Blanchard and Leigh (2013; IMF, 2012) and Fátas and Summers (2018) are extended here into a panel framework in order to assess the empirical basis of the so-called IMF "mea culpa" regarding the underestimation of Keynesian multipliers during the euro area crisis. The objections put forward by the European Central Bank, the European Commission and other authors against the "underestimation" thesis are tested and refuted. The results support the "mea culpa" and highlight that the underestimation of multipliers can concern both the short and the long term.

Brancaccio: Università del Sannio, Italy,
email: emiliano.brancaccio@unisannio.it
De Cristofaro: Scuola Superiore Sant'Anna, Italy
email: fabiana.decristofaro@gmail.com

How to cite this article:
Brancaccio E., De Cristofaro F. (2020), " Inside the IMF "mea culpa": A panel analysis on growth forecast errors and Keynesian multipliers in Europe", *PSL Quarterly Review*, 73 (294): 225-239.

DOI: https://doi.org/10.13133/2037-3643_73.294_2

JEL codes:
E23, E27, E62

Keywords:
GDP forecast errors, Keynesian multipliers, panel analysis

Journal homepage:
http: //www.pslquarterlyreview.info

The seminal work by Blanchard and Leigh (IMF, 2012; Blanchard and Leigh, 2013) has reopened a lively debate on the estimation of Keynesian multipliers and the related effects of fiscal policies. By running a cross sectional regression across 27 European economies, the authors provide evidence supporting the view that the main institutional forecasters underestimated fiscal multipliers between 2010 and 2011.

The approach adopted by Blanchard and Leigh to test their hypothesis represents an adaptation to macroeconomics of a method which has a very long tradition in finance (Timmermann and Granger, 2004). Blanchard and Leigh assume that if forecasters exploited all the information available at the time the prediction was made and used the correct model

* This article develops some analyses of the two authors presented during the seminar "Il buono e il cattivo del'austerità" dedicated to the presentation of the book *Austerity* by Alberto Alesina, Carlo Favero, and Francesco Giavazzi, held at the Bocconi University in Milan on 19 February 2019, with speeches by Emiliano Brancaccio, Veronica De Romanis, Mario Monti, the authors of the book, and moderator Ferruccio de Bortoli. We are grateful to the participants in the seminar, Carlo D'Ippoliti, Sebastian Gechert, Andrea Roventini, and two anonymous referees for their useful comments. The usual disclaimers apply.

to predict, then we should not observe any relation between growth forecast errors and the forecast of fiscal consolidation. On the contrary, if larger fiscal contractions are related to systematic changes in GDP growth forecast errors, this might mean that the effect of the fiscal multiplier has not been correctly estimated. More specifically, a negative correlation between planned fiscal consolidation and GDP forecast errors would highlight an underestimation of fiscal multipliers. Regression results show that fiscal multipliers were underestimated by one percentage point (of GDP) on average between 2010 and 2011. However, the evidence is not robust about subsequent years. Blanchard and Leigh (2014) interpret this result as a possible learning by forecasters.

Since its first publication on the International Monetary Fund's (IMF) *World Economic Outlook* of October 2012, Blanchard and Leigh's test gave rise to a broad debate about its validity and its policy implications. From the outset, their study has received criticism not only from the academic community and international institutions, but also from the major financial newspapers and other media (Giles, 2012). Among the reasons for these wide criticisms is certainly Blanchard's top position, not only in the role of chief economist of the IMF but also in the mainstream tradition of economic theory and policy. Although on several circumstances he has marked his distance from the neoliberal orthodoxy guidelines, Blanchard has nevertheless remained implanted in the theoretical and political neoclassical tradition, which is characterized by a notable trust in the spontaneous mechanisms of the market and by a very moderate use of Keynesian interventionism. In some recent contributions he has shown much less confidence in the free play of the market and has insisted on the need a Keynesian "evolution" or even a "revolution" in the guidelines of contemporary economic policy (Blanchard, 2018; Blanchard and Summers, 2019). However, his position in the mainstream theory and policy was never questioned by him, and even his latest criticisms of the concept of "natural equilibrium" do not seem to cross the boundaries of the traditional neoclassical approach (Brancaccio and Saraceno, 2017; Brancaccio and Califano, 2018; Blanchard and Brancaccio 2019; see also other contributions in Brancaccio and De Cristofaro, 2019). Also for these reasons, the exhumation of Keynesian multipliers and the related criticism of austerity policies by Blanchard must have been understood by the mainstream citadel as an implicit exhortation to recite a "mea culpa", and as an even more alarming provocation because launched from within the walls of the International Monetary Fund (Krugman, 2013).

The opposition to Blanchard and Leigh also involved the most important European institutions at the forefront. The European Commission stated that "the link between past forecast errors and planned fiscal consolidations is not robust and should not be taken as indirect proof for a larger consolidation multiplier" (European Commission, 2012); in particular, the coefficient estimated by Blanchard and Leigh would rather capture "a negative response of investors to possibly insufficient fiscal effort in countries with severe debt problems". In the same vein, the European Central Bank replied to Blanchard and Leigh's test a few months later stating that "over the longer run, fiscal consolidation has sizeable benefits, not only in terms of fiscal sustainability but also when measured in terms of GDP" and claiming that results from its New Area-Wide Model "are in line with the available empirical evidence which points to a higher degree of success for expenditure-based fiscal consolidations" (ECB, 2012). Moreover, Alesina et al. (2019) suggested that Blanchard and Leigh's results might actually be driven by high debt countries which did fiscal consolidation harsher than planned and also supported the ECB in stating that fiscal consolidation is less contractionary when based on expenditure cuts rather than tax increases. Further objections to Blanchard and

Leigh's test came from Górnicka et al. (2020), who criticized the empirical strategy for the lack of a counterfactual. More generally, all these criticisms of Blanchard and Leigh agree that, beyond their actual size, Keynesian multipliers can only have relevance in the short term.

It must be said that not all the commentators criticized Blanchard and Leigh's test: other contributions supported their thesis (Gechert et al., 2019; Fátas and Summers, 2018; House et al., 2019; see also Brancaccio and De Cristofaro, 2019).

Despite the broad debate inspired by Blanchard and Leigh's test, however, it is interesting to note that few contributions have tried to replicate and extend their exercise. Among the attempts dedicated to a development of Blanchard and Leigh's test, particularly relevant are those proposed by the European Commission (2012), Gechert et al. (2019), and Górnicka et al. (2020). While the first two contributions focus on European Commission forecasts and in some respects use a different approach, the latter one exactly replicate Blanchard and Leigh's regression and provide evidence about IMF forecasts for 2010 and 2011. A further extension of Blanchard and Leigh's test is provided by Fátas and Summers (2018), who investigate the permanent effects of fiscal consolidation using a two-stage least square regression. The first stage of Fátas and Summers' model corresponds to Blanchard and Leigh regression, whose fitted values are then used as a covariate in the second stage to explain 5-years forecast errors in potential output growth. The rationale behind this exercise is to measure to what extent the unexpected fall in GDP originated by fiscal consolidation policies might have contributed to permanently lower the path of potential output.

In this article we adopt the two stage least squares framework by Fátas and Summers, which incorporates Blanchard and Leigh's test. A novelty of our contribution is that we extend Blanchard and Leigh's and Fátas and Summers' analyses until 2018. This will allow us to take advantage of both the cross country and the time series dimension of the data; as is well known, panel data analysis provides more accurate inference of model parameters since longitudinal data contain more degrees of freedom and more sample variability than cross-sectional or time series data (Hsiao, 2007).

By adopting this methodology we intend to raise some objections to the criticisms made against Blanchard and Leigh's test and the related attempts to re-establish the validity of restrictive fiscal policies, with particular regard to cuts in public spending. First of all, we propose tests which are not affected by the problems that characterize the analyses of the European Commission (2012) and the European Central Bank (2012), such as a very small number of observations and a conjecture of an exogenous change in sovereign bond yields between 2010 and 2011, which in fact could also be considered as endogenous.[1] Second, we indirectly test the conjecture that the fiscal multiplier associated with policies on the taxation side is larger than that associated with policies on the expenditure side. Third, we test the hypothesis that highly indebted countries drive the correlation between growth forecast errors and planned fiscal consolidation. Finally, we analyse how fiscal changes are transmitted to the long run, and we test the hypothesis that forecasters learnt from their mistakes versus the alternative hypothesis of time-varying fiscal multipliers. As we shall see, our analysis will provide new evidence to support the underlying motivations for the IMF's "mea culpa" on the contractionary effects of fiscal consolidations in Europe.

[1] As pointed out in Blanchard and Leigh (2013), controlling for ex-post events which were unknown at the time of the forecast raises an endogeneity issue. Indeed, an ex-post increase in sovereign bond yields could cause lower than expected growth but could also be triggered by lower than expected growth.

The article is structured as follows. In section 1 we will describe the methodology applied and the data. Section 2 presents the baseline model results together with some robustness checks. Section 3 discusses two conjectures on fiscal consolidation made by the critics of Blanchard and Leigh's test: the first one states that spending cuts are less recessionary than tax hikes, and the second one suggests that it was the actual fiscal stance to be underestimated rather than the multipliers. Section 4 tests the forecasters' learning hypothesis and the time-varying multipliers hypothesis. Section 5 discusses the long run effect of fiscal policy. Section 6 presents results controlling for confounding factors. Section 7 concludes.

1. Empirical methodology and data

We can interpret Blanchard and Leigh's and Fátas and Summers' tests as a one stage and a two-stages exercise, respectively. We shall use Blanchard and Leigh's test for most of our argument. Then, by following Fátas and Summers, we shall use fitted values from Blanchard and Leigh's first stage to assess the long run impact of fiscal policy in a two-stages least squares (TSLS) framework.

Blanchard and Leigh's first stage is based on the following idea. Under rational expectations and assuming the "true" macroeconomic model is used for prediction, we should observe no correlation between growth forecast errors and the forecast of fiscal stance. On the contrary, if fiscal multipliers were underestimated we would observe a systematic negative correlation: for example, an increase in fiscal consolidation would imply an even more negative growth forecast error since actual growth would be smaller than the predicted one. The baseline specification of the first stage, which essentially replicates Blanchard and Leigh in a panel framework, is therefore the following:

$$\Delta GDP^{fe}_{i,t:t+1} = \alpha + \beta \Delta F_{i,t:\,t+1} + \lambda_t + \varepsilon_{i,t:t+1} \quad (1)$$

with $i = 1, \dots, 21$ indicating the European economies in our sample and $t = 2010, \dots, 2017$ representing the year of the forecast. $\Delta GDP^{fe}_{i,t:t+1}$ is the forecast error of the cumulative GDP growth over a two-year period, and $\Delta F_{i,t:\,t+1}$ is the planned fiscal consolidation over the same time span (on the exogeneity of this measure, cfr. Blanchard and Leigh, 2013; see also Fátas and Summers, 2018). Following Blanchard and Leigh, we include a vector λ_t of time fixed effects as well. Growth forecast errors are computed as the difference between actual cumulative growth over 2 years and its forecast:

$$\Delta GDP^{fe}_{i,t:t+1} = \Delta GDP_{i,t:t+1} - \Delta GDP^{f}_{i,t:t+1}$$

Forecasts data, $\Delta GDP^{f}_{i,t:t+1}$, are collected using all the *World Economic Outlook* (WEO) vintages from 2010 to 2017, while the latest release at the time of writing (April 2019) is used to derive the actual GDP realizations, $\Delta GDP_{i,t:t+1}$. The planned fiscal adjustment $\Delta F^{f}_{i,t:\,t+1}$ is given by the change in the discretionary fiscal effort (DFE hereafter) over two years, and the series are retrieved from the AMECO database (European Commission, 2013). It is worth mentioning that AMECO provides data for a limited set of European countries only; therefore, our analysis is confined to the following 21 economies: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta,

Netherlands, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden and the United Kingdom. The panel is balanced

In choosing the measure to adopt as explanatory variable in equation 1, we depart from Fátas and Summers and Blanchard and Leigh, who use the structural budget balance. DFE, indeed, has been argued to be more robust for estimating fiscal multipliers (Carnot and De Castro, 2015, and additionally it overcomes endogeneity issues in the regression with the potential output in the second stage (Gechert et al., 2019). In equation 1, β can thus be interpreted as the coefficient measuring the average fiscal multiplier underestimation between 2010 and 2017, while in the cross-section regression (by Blanchard and Leigh) it is specific of the two-year period considered.

Results of this baseline model will be discussed in section 2. We readapt this first specification and develop two extensions in order to assess two popular conjectures about fiscal multipliers. Results from the aforementioned models are discussed in section 3. A further specification that assesses two popular interpretations of fiscal multipliers' underestimation is discussed in section 5.

The second stage of the test is aimed at measuring the long-term effect of fiscal policy by regressing $t+5$ forecast errors of cumulated potential output growth on the fitted values from the first stage – that is, on the unexpected GDP change associated with the planned fiscal consolidation over a two-year interval. In other words, the effect of a cyclical policy shock on GDP is isolated here via the fitted values from the first stage, and then persistency is assessed through the impact of the unexpected GDP variation on potential output $t+5$ forecast errors. According to Fátas and Summers, this exercise addresses the issue of measuring *permanent* effects of fiscal policy changes, as "potential output can be seen as a long-term forecast for GDP" (Fátas and Summers, 2018). The idea is that if a cyclical shock is perceived as transitory, a fall in GDP should lead to a small or no change in potential output. It is worth mentioning, however, that the authors warn potential output is a constructed measure and therefore may contain noisy and limited information on long-term GDP. Nevertheless, they finally accept it as the best indicator available to estimate permanent changes in GDP patterns (on potential output and output gap measures, see e.g. Palumbo 2008, 2015). The second-stage equation is therefore the following:

$$\Delta PotGDP^{fe}_{i,t:t+5} = \alpha + \gamma \Delta \widehat{GDP^{fe}_{i,t:t+1}} + \lambda_t + \varepsilon_{i,t:t+1} \quad (2)$$

where $\Delta PotGDP^{fe}_{i,t:t+5}$ is the five-years growth forecast error of potential output, $\Delta \widehat{GDP^{fe}_{i,t:t+1}}$ are the fitted values from the first stage, and λ_t is a vector of time fixed effects. Potential output forecasts were collected from the 2010-2017 vintages of the IMF's WEO.[2] In equation 2, γ measures the persistence of changes in output caused by fiscal policy variations: in other words, the long run effect of fiscal policy. It is worth noticing that if $\gamma = 1$ the multiplier effect in the short run will be transferred one to one to the long run. This point is crucial as, according to the prevalent model, the long run can be thought of as a hypothetic situation in which the production level would be completely independent from effective demand, and hence from fiscal policy too. Blanchard and Leigh's test – and Fátas and Summers' test even more so – contribute to show that this equilibrium is in fact extremely difficult to reach as their empirical

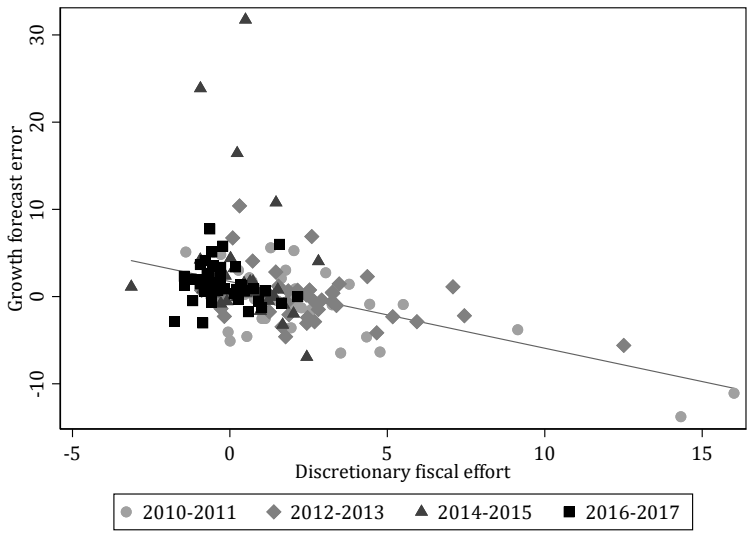
[2] Before computing forecast errors, GDP and potential output series need to be adjusted to avoid distortions in the estimations. Indeed, since October 2016 the WEO reports updated data employing the ESA2010 criteria. These changes make comparisons across vintages impossible. Therefore, we follow Fátas and Summers (2018) and adjust the series using the methodology proposed by these authors.

analysis shows some degree of persistency or even permanency of the effects of fiscal shocks on GDP (Califano and De Cristofaro, 2018).

2. Underestimation of Keynesian multipliers

As previously discussed, we estimate the baseline specification of the Blanchard and Leigh's first stage test in order to assess fiscal multipliers' underestimation. In table 1, we replicate Blanchard and Leigh's test using our panel dataset, which includes 21 countries over the period 2010-2017 for an overall sample size of 168 observations. Column 1 reports the OLS estimation results of the baseline regression described by equation 1. As suggested by Blanchard and Leigh, we apply the Newey-West procedure in order to correct standard errors for serial correlation of type MA(1) due to the use of two-year overlapping intervals.

Figure 1 – *An extension of the Blanchard and Leigh's test* (*IMF, 2012*; *Blanchard and Leigh, 2013*)



The panel estimate of β is −0.643 (t-statistic = −4.97), which is smaller in absolute terms than the coefficient found by Blanchard and Leigh for the forecasts made in early 2010 (1.095), although still strongly statistically significant. The estimated parameter suggests that the average multiplier underestimation has been of about 0.64 percentage points of GDP for the forecast period 2010-2017.[3]

Figure 1 illustrates the result using a scatter plot. As we shall further discuss in the following sections, it seems clear that the negative correlation between fiscal consolidation and

[3] As argued in Blanchard and Leigh (2013), the constant term "has no strong economic interpretation" (it can be interpreted as the sample mean of the growth forecast error minus the slope times the sample mean of the fiscal consolidation measure).

growth forecast errors is stronger for the earliest part of the sample – that is, 2010-2011 – so that a smaller β in the overall panel estimation does not come as a surprise.

Table 1 – *First-stage estimates of the baseline model, robustness tests and extensions*

	(1) Baseline	(2) Euro area	(3) Quantile	(4) Robust	(5) GT	(6) High-low debt	(7) Learning hp	(8) Time varying multip. hp a)	(9) Time varying multip. hp b)
Fiscal consolidation	−0.643***	−0.717***	−0.642***	−0.545***					−0.873***
	(0.129)	(0.128)	(0.093)	(0.081)					(0.1605)
Expenditure					1.662***				
					(0.432)				
Revenues					−0.830*				
					(0.498)				
Fiscal consolidation* high debt countries						−0.637***			
						(0.134)			
Fiscal consolidation* low debt countries						−0.713***			
						(0.257)			
Fiscal consolidation 2010-2011							−0.844***	−1.081***	
							(0.054)	(0.103)	
Fiscal consolidation 2011-2012							−0.702***	−0.937***	
							(0.218)	(0.256)	
Fiscal consolidation 2012-2013							−0.423***	−0.663***	
							(0.095)	(0.137)	
Fiscal consolidation 2013-2014							−0.150	−0.515	
							(0.290)	(0.318)	
Fiscal consolidation 2014-2015							−1.019	−1.245	
							(0.796)	(0.777)	
Fiscal consolidation 2015-2016							−1.270	−1.424	
							(1.289)	(1.144)	
Fiscal consolidation 2016-2017							−0.101	−0.152	
							(0.630)	(0.748)	
Fiscal consolidation 2017-2018							−0.480	−0.177	
							(0.665)	(0.761)	
Growth SD								1.380**	1.374***
								(0.534)	(0.525)
Constant	2.383***	2.573***		2.110***	2.776***	2.420***	2.816***	0.640	0.206
	(0.520)	(0.624)		(0.529)	(0.564)	(0.529)	(0.527)	(0.962)	(0.946)
Observations	*168*	*128*	*168*	*168*	*168*	*168*	*168*	*168*	*168*
Adj. R-squared	*0.260*	*0.267*	*0.168*	*0.249*	*0.262*	*0.260*	*0.205*	*0.411*	*0.420*

Notes: Newey-West estimates of the baseline and extended models in columns 1, 2, 5, 6, 7, 8, and 9. Quantile and robust estimates of the baseline specification in column 3 and 4, respectively. Standard errors in parenthesis, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: elaboration on WEO 2010-2017 databases, and AMECO.

Columns 2, 3, and 4 of table 1 provide some robustness checks to the baseline estimation, while columns 5 and 6 show two extensions further discussed in the next section. As a first step, equation 1 is re-estimated limiting the analysis to the euro area countries. The estimated coefficient now is slightly higher in absolute value, as shown in column 2. This result is consistent with the view that euro area countries, having been particularly strongly hit by the crisis and being bound to the common currency with monetary policy at its zero lower bound, might have experienced larger multipliers (Almunia et al., 2010; Christiano et al. 2011). In

column 3, we report a quantile estimation of the baseline model. We might be concerned about the presence of some extreme observations – which are easily detectable from the scatterplot in figure 1 – and indeed this approach is proved to be more robust than the simple OLS (differently from the OLS estimation, quantile regression minimizes the sum of absolute residuals around the median instead of the sum of squared residuals around the mean). Interestingly, the estimated coefficient in column 3 is not statistically different from the OLS one in column 2. As a further robustness check, in column 4 we report the estimate of a robust regression which down-weights observations with larger absolute residuals. The estimated coefficient in column 4 is smaller, though similar, than the one for the baseline model. Columns from 5 to 8 – which represent several extensions of the baseline specification – will be discussed in the next sections.

3. Assessing two conjectures about fiscal consolidation

Having checked the robustness of the baseline specification, we are interested in assessing two popular conjectures about fiscal multipliers.

The first one states that spending cuts are less recessionary than tax hikes (Mountford and Uhlig, 2009; Alesina and Ardagna, 2010; ECB, 2012; Alesina et al., 2019). Moreover, if that is the case, forecasters might have failed to compute multipliers for one or the other side of the fiscal balance.

The second conjecture calls into question high-debt countries as those that did more consolidation than predicted at the beginning of the period, particularly during the sovereign debt crisis (European Commission, 2012; Alesina et al., 2019). This would make the estimation of β biased due to an underestimation of the fiscal stance rather than of the Keynesian multipliers. It is a challenge for Blanchard and Leigh's interpretation.

In column 5 of table 1 we investigate whether the estimated negative correlation of the baseline may be driven by government spending or revenue changes. Using DFE as the policy variable instead of structural balance makes the job easier compared to Blanchard and Leigh: the AMECO dataset separately reports estimates of the expenditure and the revenue side of the fiscal effort, while Blanchard and Leigh had to make several assumptions in order to decompose the structural balance (Blanchard and Leigh, 2013). Our results show that fiscal multipliers were underestimated for both sides of the fiscal balance, and especially for the spending side. The coefficient associated to government expenditure is in fact 1.66, which is significantly higher in absolute terms than the coefficient estimated for the revenue side, equal to –0.83 and less statistically significant. This result reinforces the finding by Blanchard and Leigh – whose coefficients are economically meaningful and statistically significant though not statistically different from each other – and that by Gechert et al. (2019), who obtain non-significant coefficients.

Interestingly, our results seem to suggest that contractionary effects were underestimated especially in case of spending cuts, therefore implying a larger multiplicative effect. This might be in contrast with the evidence suggested by the ECB (2012) and Alesina et al. (2019), who find that "spending cuts have been associated with very small downturns" while "plans based on tax increases are associated with large recessions" (ibid., p. 196).

In column 6, we analyse the second conjecture and extend the baseline model by splitting the estimated coefficient in two parts. We thus allow the estimation of two different slopes:

one for high-debt countries and one for low-debt countries. We adopt one of the Maastricht convergence criteria as a threshold and define high-debt countries as those exhibiting a public debt to GDP ratio larger than 60%. The estimation in column 6 suggests that the correlation found in the baseline specification is not driven by a specific group of countries and holds for both high- and low-debt countries. Specifically, this result seems to be at odds with the idea – supported by Alesina et al. (2019) among others – that Blanchard and Leigh's findings are due to unexpected GDP losses experienced by some countries because of their high indebtment. It is worth mentioning that we replicated the previous estimation with a different threshold, set at 100%, and the previous results hold as well: the estimated coefficients are still both statistically significant and not statistically different one from the other.

4. Forecasters learning vs. time varying fiscal multipliers

Blanchard and Leigh's test – represented here by the first stage estimation – suggest that forecasters failed to predict GDP growth because of underestimation of fiscal multipliers. However, the detected negative correlation between planned fiscal consolidation and growth forecast errors broke up in recent years. Two explanations have been proposed on that regard: forecasters learnt from their mistakes, or multipliers are getting smaller in more recent times (Auerbach and Gorodnichenko, 2012, 2013; Blanchard and Leigh, 2014; Górnicka et al., 2020).

In order to evaluate a possible learning effect by forecasters, in our TSLS framework we keep the panel structure and let the coefficient β to vary over time. Interestingly, we find that, differently from Blanchard and Leigh (2014), the coefficient of interest at the first stage is statistically significant until the two-year period 2012-2013, whereas they find statistical significance only for the two year-period 2010-2011. According to the results shown in column 7 of table 1, we can say that if there was a learning effect it started from 2013. Interestingly, as we shall see in section 6 below, once we control for other factors influencing the investigated relationships it emerges that fiscal multipliers were underestimated in the 2014-2015 two year-period too.

With respect to Blanchard and Leigh's learning hypothesis, a possible alternative interpretation relies on the literature on time-varying fiscal multipliers (Auerbach and Gorodnichenko, 2012, 2013). The idea is the following. It could be argued that forecaster failed to predict GDP growth because fiscal multipliers changed over time if, after controlling for business cycle variations, the coefficient attached to fiscal consolidation becomes statistically insignificant. In this section we propose a simple exercise for discerning which hypothesis may better justify the absence of correlation between the fiscal stance and growth forecast errors in the recent years.

As regards learning, we see that fiscal multipliers might have been underestimated from 2010-2011 until 2012-2013, and in the 2014-2015 two year-period as well: such a 'hole' in statistical significance seems to raise doubts on a possible learning by forecasters. As regards time varying multipliers, we used the standard deviation of GDP growth in order to capture business cycle variability. The more different business cycle phases take turn, the more we expect growth variability being higher and the more it must be related with larger forecast errors. In our exercise, it turns out that growth variability explains growth forecast errors but fiscal stance does it as well. This happens in both the baseline specification (column 9) and in the model with time-varying parameters (column 8). It is worth noticing that, once we take

into account business cycle volatility, the 2013-2014 and 2014-2015 periods are just short of being statistical insignificant at the conventional threshold (p-value around 0.11). As we shall see in section 6, these results are confirmed after controlling for the usual set of variables. Interestingly, the impact of growth variability on growth forecast errors is positive, meaning that the more a country's growth is volatile the more forecasters are conservative in their estimations, predicting lower GDP growth than the actual realization.

5. On the long-term effects of fiscal policy

In table 2 we report the TSLS estimation of equation 2.[4] The baseline model in column 1 essentially replicates Fátas and Summers' regression, while columns from 2 to 7 show the second stage of the robustness checks, and extensions discussed in the previous sections. It is worth pointing out that the coefficient of the second-stage regression is approximately 1 for every specification implying that the real effects of fiscal policy on GDP growth are entirely transmitted to the long run. The unexpected growth variation associated to changes in fiscal policy is shown to have an impact on the potential output growth for 5 years ahead growth.

Table 2 – *Second-stage estimates of the baseline model, robustness tests and extensions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Euro area	GT	High-low debt	Learning hp	Time-varying multip. hp a)	Time-varying multip. hp b)
Fitted growth	1.033***	1.183***	1.051***	1.045***	0.973***	1.463***	1.612***
forecast errors	(0.335)	(0.346)	(0.322)	(0.327)	(0.241)	(0.185)	(0.264)
Growth SD						2.216***	2.218***
						(0.212)	(0.395)
Constant	–1.682	–0.848	–1.700	–1.694	–1.622	–6.359***	–6.577***
	(1.624)	(1.934)	(1.658)	(1.605)	(1.610)	(1.021)	(1.087)
Observations	*168*	*128*	*168*	*168*	*168*	*168*	*168*
R-squared	*0.540*	*0.579*	*0.545*	*0.544*	*0.524*	*0.760*	*0.748*

Notes: Two-Stages Least Squares estimates (second stage only) of the baseline and extended models. Standard errors in parenthesis, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: elaboration on WEO 2010-2017 databases, and AMECO.

The results are robust also when directly estimating the following regression:

$$\Delta PotGDP^{fe}_{i,t:t+5} = \alpha + \gamma \Delta F_{i,t:\,t+1} + \lambda_t + \varepsilon_{i,t:t+1}$$

Interestingly, the estimated coefficient is similar to the one we get at the first stage for the baseline model (–0.664, with t-statistic = –3.42), confirming the result that real effects of fiscal policy are transmitted one to one to the long run. In this respect out findings differ from those by Gechert et al. (2019) who do not get a statistically significant coefficient, even though the coefficient estimated by them is large and with the correct sign.

[4] TSLS estimation of equation 1 yields the same coefficients shown in table 1 but with slightly smaller standard errors as we do not implement the Newey-West correction. As we wanted to be conservative, we decided to show robust estimates in table 1.

6. Additional control variables

In this section we discuss the robustness of our TSLS estimation to potential omitted variable bias, that is to possible factors that might have determined both fiscal consolidation and lower than expected growth. It is worth mentioning that it only makes sense controlling for factors that at the time of the forecast were in the information set of the forecasters. Controlling for ex-post developments that were not in the forecasters' information set is not allowed and may expose to reverse causality issues. As an example, according to Alesina et al. (2019) the spike in the interest rates that took place in 2010 and spurred the eurozone crisis might have led several (high-debt) countries to try to gain back the markets' confidence by implementing harsher austerity measures than planned. The growth forecast errors, thus, would not have been caused by an underestimation of the Keynesian multiplicative effect but by an underestimation of the fiscal contraction actually realized in such turbulent stage of the European crisis. However, it goes without saying that an ex-post rise in sovereign borrowing costs might have been caused by lower than expected growth as well as it might have caused itself lower growth (Romer and Romer, 2010; Cottarelli and Jaramillo, 2012Brancaccio and De Cristofaro, 2019a). Controlling for such a factor would not help to disentangle the effect the authors want to capture and would make the economic interpretation of β unclear.

We follow Blanchard and Leigh in the choice of the control variables. One of the criticisms that could be raised to our exercise is that lower than expected growth might have been caused by sovereign debt problems rather than fiscal consolidation. In order to control for such an effect, we check whether our baseline model is robust to the introduction of the initial debt to GDP ratio and the initial fiscal balance to GDP ratio. We take care of using the values from the corresponding WEO vintage each year in order to ensure that the information was actually in the information set of the IMF forecasters. Next, we want to check whether lower than expected growth might have been determined by financial market stress. In order to control for such a mechanism, we used the recently updated dataset by Laeven and Valencia (2018 which classifies systemic banking crises from 1970 to 2017. Based on these data, we built a dummy variable taking on value 1 every time a country registers a systemic banking crisis at least in one year in the two-year-period that we use in the analysis. Further, we assessed the possibility that forecasters made bad predictions because they may have failed to adjust to news about the state of the economy, implying serial correlation in forecast errors (Coibion and Gorodnichenko, 2012). In order to account for this possible correlation, we used the initial GDP growth forecast and the initial potential GDP growth forecast. Finally, we controlled for a possible effect of external imbalances that might have determined lower than expected growth. We follow Blanchard and Leigh and use lagged values in $t - 4$ taking into account pre-crisis imbalances.[5] As a measure of imbalance, we adopt the current account to GDP ratio as reported by the WEO, and the net financial position to GDP ratio as computed by Lane and Milesi-Ferretti (2015).

[5] The results hold as well if we use lagged values in $t - 1$.

Table 3 – *TSLS estimates of the baseline model and extensions*

	(1) Baseline	(2) G/T	(3) High-low debt	(4) Learing hp	(5) Time varying multip. hp a)	(6) Time varying multip. hp b)
First stage						
Dependant var.: growth forecast errors						
Fiscal consolidation	−1.063***					−0.946***
	(0.293)					(0.268)
Expenditure		2.020***				
		(0.569)				
Revenues		−1.245				
		(0.781)				
Fiscal consolidation high debt countries			−1.071***			
			(0.308)			
Fiscal consolidation low debt countries			−1.029***			
			(0.300)			
Fiscal consolidation 2010-2011				−1.145***	−1.165***	
				(0.196)	(0.176)	
Fiscal consolidation 2011-2012				−0.940***	−1.082***	
				(0.213)	(0.228)	
Fiscal consolidation 2012-2013				−0.671***	−0.743***	
				(0.216)	(0.191)	
Fiscal consolidation 2013-2014				−0.485	−0.535	
				(0.367)	(0.358)	
Fiscal consolidation 2014-2015				−1.404**	−1.267*	
				(0.674)	(0.678)	
Fiscal consolidation 2015-2016				−1.209	−1.138	
				(1.025)	(0.899)	
Fiscal consolidation 2016-2017				−0.443	−0.015	
				(0.562)	(0.600)	
Fiscal consolidation 2017-2018				−0.239	0.078	
				(0.554)	(0.707)	
Growth SD					1.435***	1.426***
					(0.275)	(0.268)
Constant				−0.323	1.049	0.570
				(2.552)	(1.934)	(1.977)
Second stage						
Dependant var: potential output growth forecast errors						
Fitted growth forecast errors	1.925***	1.839***	1.920***	1.632***	1.616***	1.746***
	(0.414)	(0.437)	(0.411)	(0.313)	(0.316)	(0.395)
Growth SD					2.518***	2.321***
					(0.409)	(0.426)
Constant	−8.035***	−8.113***	−8.040***	−8.743***	−7.827***	−7.747***
	(2.563)	(2.707)	(2.569)	(2.471)	(1.440)	(1.531)
Controls	yes	yes	yes	yes	yes	yes
Observations	*168*	*168*	*168*	*168*	*168*	*168*
R-squared	*0.688*	*0.693*	*0.688*	*0.697*	*0.801*	*0.786*

Notes: Two-Stages Least Squares estimates of the baseline and extended models. Standard errors in parenthesis, *** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$.
Source: elaboration on WEO 2010-2017 databases, and AMECO.

Table 3 shows the TSLS estimation results once we control for the aforementioned variables. For all specifications we find larger coefficients in absolute values for our main variables of interest, indicating that in the estimations of tables 1 and 2 the parameter of interest was positively biased in the first stage and negatively in the second.

Column 2 of table 3 adds the control variables to the specification that separately estimates the correlation of government spending and taxation with growth forecast errors. The estimation in column 2 suggests that, differently from the specification without controls, fiscal multipliers were underestimated especially on the expenditure side, with the coefficient for revenues now not statistically significant. This piece of evidence can be interpreted once again as not supporting the idea that spending cuts are less recessionary than tax hikes.

Column 3 shows that even controlling for the abovementioned confounding mechanisms, fiscal multipliers were underestimated for both low-debt countries and high-debt countries. Interestingly, the coefficients of interest are both highly statistically significant and with similar size. It is worth noting, in columns 4 and 5, that after controlling for the effect of other confounding factors, the underestimation of fiscal multipliers becomes statistically significant also in the 2014-2015 two-year period, even when business cycle volatility is considered.

Finally, column 6 shows that fiscal multipliers underestimation is statistically significant and essentially of the same magnitude when controlling for cycle variability too. Interestingly, TSLS estimation of the second-stage suggest that cyclical variations in fiscal policy are transmitted to the long run almost doubled.

7. Conclusions

This article replicates and extends the Blanchard and Leigh (IMF, 2012; Blanchard and Leigh, 2013) and Fátas and Summers (2018) tests, shows that their findings are confirmed in a panel framework and highlights some weaknesses in the arguments of those who criticized the results of these tests. First, we find that over the period 2010-2018 fiscal multipliers were underestimated by about 0.64 percentage points of GDP and the consequent unexpected GDP growth loss was transmitted one to one in the long run. Second, the empirical evidence contrasts the argument that the expenditure multiplier is smaller than the tax multiplier. Third, contrary to the view that highly indebted governments implemented harsher fiscal consolidation measures determining lower growth than expected, our analysis shows that fiscal multipliers were underestimated for both high-debt and low-debt countries. Finally, this paper provides evidence that fiscal multipliers might have been underestimated from 2010-2011 until 2012-2013 and in the 2014-2015 two year-period as well: such a 'hole' in the statistical significance seems to raise doubts on possible learning by forecasters, and suggests testing alternative interpretations.

Proposed by Blanchard and Leigh and extended by Fátas and Summers, the theses according to which growth forecast errors might have been determined by an underestimation of fiscal multipliers seem to resist the refutation attempts. Furthermore, these underestimates seem to persist over time and are not subject to learning processes by institutional forecasters and consequent revisions. These results give further support to the criticisms of European austerity policies, with particular regard to those based on cuts in public spending. The implications appear relevant with regard not only to the historical interpretation of the

eurozone crisis of 2010-2012 but also to the crisis triggered by the coronavirus and the economic policies that will be adopted in the years to come.

It should be noted that some of the results reported here are related to standard hypotheses of the mainstream theoretical approach. In particular, in determining potential output we follow a criterion suggested by Fátas and Summers (2018). These sections of the analysis can therefore be understood as an 'internal critique' of the prevailing approach to economic theory and policy. Future research may be devoted to a replication of these tests in the absence of any reference to mainstream theoretical hypotheses, in particular by adopting alternative definitions for potential production and the output gap (Palumbo, 2008, 2015).

In conclusion, the IMF "mea culpa" signed by Blanchard and Leigh seems well founded: the conjecture that the IMF and the other forecasters failed to predict GDP growth – and may fail to predict it in the future – because of an 'anti-Keynesian prejudice' is still in place. The results of the analysis, therefore, this time are 'pro-Blanchard'. It remains to be asked whether and to what extent his reversal of position has come late, possibly also due to the basic contrasts between the Keynesian paradigm and the mainstream neoclassical approach of Blanchard and the authors examined here (on this point see, among many others: Pasinetti, 2005; Roncaglia, 2011; Brancaccio and Califano, 2018; Blanchard and Brancaccio, 2019). Any evaluation on this matter is obviously open to discussion.

References

Alesina A. e Ardagna S. (2010), "Large Changes in Fiscal Policy: Taxes versus Spending", in Brown J.R. (ed.), *Tax Policy and the Economy. Volume 24* (pp. 35-68), Chicago: University of Chicago Press and National Bureau of Economic Research.

Alesina A., Giavazzi G. and Favero C. (2019), *Austerity. When it Works and When it Does Not*, Princeton: Princeton University Press.

Almunia M., Bénétrix A., Eichengreen B., O'Rourke K.H. and Rua G. (2010), "From Great Depression to Great Credit Crisis: Similarities, Differences and Lessons", *Economic Policy*, 25 (62), pp. 219-265.

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

Auerbach A. and Gorodnichenko Y. (2013), "Output Spillovers from Fiscal Policy, *American Economic Review*, 103 (3): 141-46.

Blanchard O.J. (2018), "Should We Reject the Natural Rate Hypothesis?", *Journal of Economic Perspectives*, 31 (1), pp. 97-120.

Blanchard O.J. and Brancaccio E. (2019), "Crisis and Revolution in Economic Theory and Policy: A Debate", *Review of Political Economy*, 31 (2), pp. 271-287.

Blanchard O.J. and Leigh D. (2013), "Growth Forecast Errors and Fiscal Multipliers", *American Economic Review: Papers and Proceedings*, 103 (3), pp. 117-120.

Blanchard O.J. and Leigh D. (2014), "Learning about Fiscal Multipliers from Growth Forecast Errors," *IMF Economic Review*, 62 (2), pp. 179-212.

Blanchard O.J. and Summers L. (2019), "Ripensare le politiche macroeconomiche: evoluzione o rivoluzione?", *Moneta e Credito*, 72 (287), pp. 171-195.

Brancaccio E. and Califano A. (2018), *Anti-Blanchard Macroeconomics. A Comparative Approach*, Cheltenham (UK) and Northampton (MA): Edward Elgar.

Brancaccio E. and De Cristofaro F. (2019a), Growth forecast errors during the euro crisis: what was the cause?, *economiaepolitica.it*, 11 (17), pp. 1-4.

Brancaccio E. and De Cristofaro F. (eds.) (2019b), "Crisi e rivoluzioni della teoria e della politica economica: un simposio", *Moneta e Credito*, 72 (287), special issue.

Brancaccio E. and Saraceno F. (2017), "Evolutions and Contradictions in Mainstream Macroeconomics: The Case of Olivier Blanchard", *Review of Political Economy*, 29 (3), pp. 345-359.

Califano A. and De Cristofaro F. (2018), "Appendix 2: Contractionary Fiscal Policy and Fiscal Multipliers", in Brancaccio E. and Califano A. (eds.), *Anti Blanchard Macroeconomics. A Comparative Approach*, Cheltenham (UK) and Northampton (MA): Edward Elgar.

Carnot N. and de Castro F. (2015), "The Discretionary Fiscal Effort: An Assessment of Fiscal Policy and Its Output Effect", *European Economy – Economic Papers*, no. 543, Brussels: European Commission, Directorate-General for Economic and Financial Affairs, available at https://ec.europa.eu/economy_finance/publications /economic_paper/2015/pdf/ecp543_en.pdf.

Christiano L., Eichenbaum M. and Rebelo S. (2011), "When Is the Government Spending Multiplier Large?", *Journal of Political Economy*, 119, pp. 78-121.

Coibion O. and Gorodnichenko Y. (2012), "What Can Survey Forecasts Tell Us about Information Rigidities?", *Journal of Political Economy*, 120 (1), pp. 116-159.

Cottarelli C. and Jaramillo L. (2012), "Walking Hand in Hand: Fiscal Policy and Growth in Advanced Economies", *IMF Working Paper*, no. 12/137, Washington (DC): International Monetary Fund.

ECB – European Central Bank (2012), *Economic Bulletin*, December, Frankfurt Am Main: European Central Bank.

European Commission (2012), "European Economic Forecast. Autumn", *European Economy*, n. 7/2012, Brussels: European Commission, Directorate-General for Economic and Financial Affairs.

European Commission, (2013), "Report on Public finances in EMU", *European Economy*, n. 4/2013, Brussels: European Commission, Directorate-General for Economic and Financial Affairs.

Fátas A. and Summers L.H. (2018), "The Permanent Effects of Fiscal Consolidations", *Journal of International Economics*, 112, pp. 238-250.

Gechert S., Horn G.A. and Paetz C. (2019), "Long-term Effects of Fiscal Stimulus and Austerity in Europe", *Oxford Bulletin of Economics and Statistics*, 81 (3), pp. 647-666.

Giles, C. (2012), Robustness of IMF data scrutinised. *Financial Time*s, October 13-14.

Górnicka L., Kamps C., Koster G. and Leiner-Killinger N. (2020), "Learning about Fiscal Multipliers during the European Sovereign Debt Crisis: Evidence from a Quasi-Natural Experiment", *Economic Policy*, May, DOI: 10.1093/epolic/eiaa008.

House C.L., Proebsting C. and Tesar L.L. (2019), "Austerity in the Aftermath of the Great Recession", *Journal of Monetary Economics*, available online at DOI: 10.1016/j.jmoneco.2019.05.004.

Hsiao C. (2007), "Panel data analysis—Advantages and challenges", *Test*, 16, pp. 1-22.

IMF – International Monetary Fund (2012), "Are We Underestimating Short Term Fiscal Multipliers?", *World Economic Outlook*, box 1.1, p. 41, October; extended in Blanchard O.J. and Leigh D. (2013), "Growth Forecast Errors and Fiscal Multipliers", *IMF Working Paper*, no. 13/1, Washington (DC): International Monetary Fund.

Krugman, P. (2013), "How the Case for Austerity Has Crumbled", *New Yourk Review of Books*, June 6.

Laeven L. and Valencia F. (2018), "Systemic Banking Crises Revised", *IMF Working Paper*, n. 18/206, Washington (DC): International Monetary Fund.

Lane P.R. and Milesi-Ferretti G.M. (2018), "The External Wealth of Nations Revisited: International Financial Integration in the Aftermath of the Global Financial Crisis", *IMF Economic Review*, 66 (1), pp. 189-222.

Mountford A. and Uhlig H. (2009), "What are the Effects of Fiscal Policy Shocks?", *Journal of Applied Econometrics*, 24 (6), pp. 960-992

Palumbo A. (2008), "I metodi di stima del Pil potenziale tra fondamenti di teoria economica e contenuto empirico", *Departmental Working Papers of Economics*, no. 92, Rome: University Roma Tre.

Palumbo A. (2015), "Studying Growth in the Modern Classical Approach: Theoretical and Empirical Implications for the Analysis of Potential Output", *Review of Political Economy*, 27 (3), pp. 282-307.

Pasinetti (2005), "How Much of John Maynard Keynes Can We Find in Franco Modigliani?", *BNL Quarterly Review*, 58 (233-234), pp. 21-39.

Romer C. and Romer D. (2010), "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks", *American Economic Review*, 100 (3), pp. 763-801.

Roncaglia A. (2011), "Macroeconomics in Crisis and Macroeconomics in Recovery", *PSL Quarterly Review*, vol. 64 n. 257, pp. 167-185.

Timmermann A. and Granger C.W.J. (2004), "Efficient Market Hypothesis and Forecasting", *International Journal of Forecasting*, 20(1), pp. 15–27.