FISCAL MULTIPLIERS

Arbresh MAMUDI, State University of Tetovo,

Geoff PUGH, Staffordshire University Business School

ABSTRACT

The prevalent view on the stabilization role of fiscal policy has evolved over time. The theory on the effectiveness of fiscal policy on output stabilization is controversial and indeterminate. The empirical evidence is huge reporting heterogeneous results: the range of empirical findings varies from negative to higher than one.We apply MRA analysis to a set of 65 empirical studies on fiscal policy estimated by either single equation approaches (SEE hereinafter) or Vector autoregression (VAR) models. In order to identify the systematic influence of certain country/structural characteristics and study characteristics on estimates of the multiplier, we classify studies with respect to the model class, type of fiscal impulses, method of calculation of multipliers, time-series properties of the data, and several other control variables. In addition, we augment the MRA model with primary data accounting for structural characteristics of the countries investigated by this literature. A novel approach of this study will be to augment MRA with primary data on labour market variables to see if there is any systematic variation to be attributed to labour market institutions.

Our main findings are as follows. Fiscal multipliers differ in terms of model classes. VAR models yield significantly higher multipliers than do SEE approaches. In line with Keynesian theory, expenditure multipliers are higher than tax shock multipliers, with public consumption being the most effective shock across several specifications of our model. Permanent shocks yield higher multipliers and multipliers estimated for a longer horizon are higher than impact multipliers. Quarterly data yield lower multipliers compared to annual data. Fiscal policy in transition countries appears to be more effective than in advanced economies, contrary to a priori expectations, although the results are not stable across different specifications. In VAR analysis, different identification strategies give rise to significantly different multipliers. In line with theoretical predictions, the reported multipliers are smaller if estimated for open economies and economies in expansion and bigger for relatively more closed economies and economies experiencing a financial crisis. Structural characteristics from primary data appear to explain less of the variation of multiplier estimates; the openness channel seems to be an important determinant of the multipliers while, from the labour market characteristics, the replacement rate appears to significantly affect the value of the multiplier. The results on the publication selection test suggest that the literature is affected by publication bias.

In conclusion, our findings suggest that the heterogeneity of the reported multipliers mostly arises from the study characteristics, however, the degree of openness and labour market characteristics appear to be important factors that should be considered while estimating fiscal multipliers.

Keywords: fiscal multipliers, fiscal policy, meta-analysis

Introduction

The prevalent view on the stabilization role of fiscal policy has evolved over time. The recent crisis and the zero bound for interest rates renewed interest in the importance of fiscal policy in fighting the recession. The empirical literature on the size of the fiscal multipliers, in the post crisis period, is growing fast. However, the results are far from consensus. Estimates of multipliers are all over the map, the range of empirical findings varies from negative to higher than one. A critical evaluation on the development of the theory of fiscal policy and employment suggests that the effectiveness of fiscal policy with respect to output stabilization is controversial and indeterminate. The inconclusive theory and a variety of empirically estimated fiscal multipliers make the application of meta-regression analysis (MRA hereinafter) a suitable set of tools to review the literature, to explain the heterogeneity of the empirical results and to investigate possible publication selection bias in the literature.

We apply MRA analysis to a set of 65 empirical studies on fiscal policy estimated by a single equation approach (SEE hereinafter) and Vector autoregression (VAR) models. In order to identify the systematic influence of certain country/structural characteristic and study characteristic on estimates of the multiplier, we classify studies with respect to the model class, type of fiscal impulses, method of calculation of multipliers, time-series properties of the data, and several other control variables. In addition, we augment the MRA model with primary data accounting for structural characteristics of the countries. A novel approach of this study will be to augment MRA with primary data on labour market variables to see if there is any systematic variation to be attributed to labour market institutions.

The results are consistent with theoretical predictions. Fiscal multipliers differ in terms of model classes. VAR models yield significantly higher multipliers than do SEE approaches. In line with Keynesian theory, expenditure multipliers are higher than tax shocks multipliers, with public consumption being the most effective shock across several specifications of our model. Data characteristics are also significant in explaining the variation of estimated multipliers. Multipliers calculated on quarterly data are lower and the longer is the horizon of calculation the higher is the multiplier value. The duration of the fiscal shock also proves to be a significant determinant of multiplier values suggesting that permanent shocks yield higher multipliers. The reported multipliers in transition countries appear to be higher compared to advanced economies, contrary to a priori expectations; however, the results are not stable across different specifications. Studies controlling for structural characteristics of the economy report significantly different multipliers than conventional studies (which don't control for these characteristics). As expected, the reported multipliers are smaller if estimated for open economies and economies in expansion and bigger for relatively more closed economies and economies experiencing a financial crisis. The subsample analysis of VAR models suggest that the differences in identification strategies are significant, while different types of VAR models are insignificant factors for determining the value of multipliers. Regarding the labour market characteristics, only replacement rates appear to be a significant determinant of the multiplier effect. In addition, although the degree of openness is an important factor, contrary to theoretical predictions, variables accounting for the indebtedness of the economy, financial development and monetary policy appear to be insignificant factors in determining multipliers.

Investigating publication selection bias, the funnel plot analysis as well as the formal FAT-PET test, which are applied by using sample size as a proxy for precision, suggests that

the literature on fiscal multipliers is infected by publication bias. Controlling for publication bias is thus an important feature of this MRA.

In summary, our findings suggest that the heterogeneity of the reported multipliers mostly arises from the study characteristics: data settings and methods of estimation, while the effect of structural characteristics is mixed. However, the degree of openness and labour market characteristics appear to be important factors that should be considered while estimating fiscal multipliers.

The rest of this paper is organized as follows: In the first section a narrative review of the fiscal multipliers literature is provided. The next section describes the preparation for MRA, the data collection process and some descriptive statistics of the variables. Section 3 provides an introduction to MRA methods and the results of the analysis are provided in Section 4 along with some robustness checks. The end of the chapter provides conclusions from the MRA.

2.1 Literature review

2.1.1 Theoretical overview

Theoretical predictions about the transmission mechanism of fiscal policy in an economy are controversial and differ not only about the size of the effects of fiscal policy on macroeconomic variables but also about the direction of these effects. While most theoretical models agree on the positive effect of an expansionary fiscal shock on output, the disagreement is about crowding in vs. crowding out effects on private consumption, investment and net exports. The effectiveness of fiscal policy as a stabilization tool depends on the assumptions about the type of the economy, type of agents, the degree of openness and the exchange rate regime of the economy.

In a new classical model, with flexible prices, assuming the economy is in equilibrium and there are no spare or non-used capacities there is no role for fiscal policy. In a Keynesian model, assuming a demand deficient economy with sticky prices, and non-forward looking agents a stimulus to demand can have multiplier effects. Higher government spending increases income, employment and consumption. Aggregate demand determines output, hence, fiscal expansion causes higher output. Fiscal policy is expected to be more effective in a closed economy compared to an open economy. In a closed economy model a fiscal expansion boosts private consumption leading to an increase in aggregate demand, output and money demand, hence, if money supply is fixed, this results in an increase of interest rates, which in turns partially crowds out private investment. Consequently, output, total investment and consumption increase. In the Keynesian model, the use of fiscal stimulus, in an open economy under a flexible exchange rate regime, will increase real interest rates and, given price stickiness, the real exchange rate will appreciate, which will lead to a loss of international competitiveness and trade balance deterioration as a result. The expansionary effect of fiscal stimulus may be entirely offset by the reduction in net exports; hence there is no fiscal multiplier. In contrast, under a fixed exchange regime, monetary policy will react in order to prevent the interest rate increase, hence exchange rate appreciation, therefore, amplifying the effect of the fiscal multiplier. In a new classical view, a fiscal shock does not change international relative prices in the absence of home bias; hence the real exchange rate is not affected by the shock. If home bias is assumed, the relative price of domestically produced goods in terms of foreign produced goods increases, which leads to an appreciation of the real exchange rate (Ravn et al., 2007).

The type of agent assumption is crucial in analysing the effectiveness of fiscal policy. Most of the studies employ DSGE models, which incorporates forward looking agents and

rational expectations. RBC models, assuming flexible prices and perfect competition suggest that fiscal policy is ineffective in stimulating the economy. Their argument against the effectiveness of fiscal policy is 'Ricardian equivalence' (Barro, 1979) based on the permanent income hypothesis, the rational expectation assumption and non-liquidity constrained consumers. This hypothesis states that debt-financed government spending may have no impact on consumer spending, because the public will save the tax cut in order to pay for future tax increases that will be initiated to pay off the debt. New classical economists also argue that higher spending financed by lump-sum taxes induces a negative wealth effect, leading to a decrease in private consumption, a contemporaneous increase in labour supply and, therefore, a decrease in the marginal productivity of labour and in real wages, which results in higher output and employment. However, the multiplier is less than one due to the 'crowding out' effect of private consumption (Baxter & King, 1993). A simple NK- DSGE model, adding microfoundations and some market failure in the models of general equilibrium analysis of inter-temporal optimisation by rational economic agents, misses the Keynesian positive effect of fiscal expansion on consumption. Its predictions are in line with those of a RBC model: an increase in output and decrease in consumption, but the model predicts an increase in the real wage. However, several modification in DSGE by altering preferences, adopting non-separable utility or deep habits in consumption, introducing non-Ricardian consumers and allowing for spending reversals may provide different predictions about the effectiveness of fiscal policy. If a model adopts a non-separable utility function in consumption and leisure, the negative wealth effect of the fiscal expansion raises hours worked, which decreases leisure, therefore, the marginal utility of consumption increases. Hence, households will work more and consume more mitigating the negative wealth effect, leading to an increase in output, employment and consumption. Also, introducing habit persistence in consumption in a model where monopolistic competition is assumed in the goods market, will result in predicting no crowding out effect of fiscal policy in consumption. A spending shock will stimulate aggregate demand and labour supply and firms will respond by reducing the mark-ups and increasing labour demand. The increase in labour demand offsets the increase in labour supply and real wages rise, therefore the value of leisure in terms of consumption declines. Consequently, output and consumption increases. In a NK-DSGE model introducing non-Ricardian households who are liquidity constrained and consume all of their current income will also generate a positive effect of fiscal shock on consumption. Additionally, the monetary policy reaction is a crucial factor determining the efficiency of fiscal policy. If the monetary policy is non-accommodative or in extreme case of a zero lower bound, the monetary authorities will not react to the increased inflationary pressures, caused by positive fiscal shock, by increasing interest rates, hence the crowding out effect in private investments is mitigated. Further, if finite horizon consumers are assumed and also anticipating spending reversal as a response to increasing public debt, in a model with nominal rigidities a fiscal shock increases output and consumption. This is consistent with 'expansionary fiscal contraction', i.e. the hypothesis that an economy may in the short term grow as a consequence of fiscal austerity. Briotti (2005) suggests that the short term *expansionary* effects may occur only if fiscal austerity improves economic agents' expectations about their future wealth and income or in the cases when fiscal austerity improves labour market efficiency and the competitiveness of the economy.

The sign and size of the reported multipliers by theoretical studies is sensitive to the model assumptions and the range of estimates varies from negative to higher than one.For instance, Real Business Cycle (RBC hereinafter) allowing for productivity enhancing effects of public spending, report multiplier values higher than one (Linnemann 2006; Mazraani 2010).On the other hand, RBC studies accounting for distortional effects of taxation report negative multipliers (Ardagna, 2001). New Keynesian DSGE models (NK-DSGE hereinafter)

report higher multiplier values than RBC studies and the magnitude of multiplier effect in these models depends on the reaction function of the monetary authority.

Studies in the pre-crisis context usually find lower multipliers than do more recent studies. Post-crisis studies of fiscal multipliers, introducing a share of non-Ricardian consumers (Cwik &Wieland 2011), or a central bank that operates at the zero lower bound (Woodford 2011; Rendahl, 2012) report multipliers values higher than one. On the other hand, in line with the 'expansionary fiscal contraction' argument, the multiplier may be negative when including so-called non-Keynesian effects due to distortionary taxation, a wage-level increasing effect of public employment, or risk premium on interest rates for high government debt (Briotti, 2005).

Macro-econometric models (MACRO hereinafter), which are in essence extensions of the Keynesian IS-LM model combining Keynesian reactions in the short-run with neoclassical features in the long run, are still used for policy consulting and report multipliers larger than one, due to crowding in of private consumption or investment.

.

Most of the literature on the effect of fiscal policy focuses on the effect on output, while the literature exploring the effect on unemployment is much more limited. The effect on unemployment is linked to the effect on output, since an increase in output is expected to be associated with higher employment. There is a consensus among economists that the response of the labour market variables to fiscal shocks depends on the type of shock considered and the way the increases are financed (Baxter&King, 1993; Pappa, 2009; Gomes, 2009; Monacelli et al., 2010). Baxter and King (1993) suggest that if the government spending is financed by lump sum taxes, this will cause an increase in hours worked and real wages will decrease. However, if distortionary income taxes are used to finance government spending, both hours worked and after-tax real wages will decrease. They argue that since higher taxes imply lower after-tax factor rewards, there is a strong incentive to substitute work effort with leisure and also to curtail investment, which results in lower output and employment. A fiscal stimulus affects the labour market through the 'negative wealth effect', 'wage and/or interest rate pressure' and the 'productivity enhancing' channels. The employment consequences depend on the state of the economy and on which of these channels prevail. In a new classical economy, a positive fiscal shock, financed by a lump-sum tax, induces the 'negative wealth effect' and an increase in the labour supply. However, the increase in labour supply may be entirely offset by the decrease of private employment due to the 'wage and/or interest rate pressure', since a fiscal shock absorbs resources and the 'productivity enhancing' channel is absent[1]. In a Keynesian economy, with underemployed resources, the 'wage and/or interest rate pressure' channel is absent. Keynesian models of both traditional partial equilibrium and new general equilibrium types, suggest that a fiscal stimulus, apart from increasing the labour supply due to the 'negative wealth effect', will also increase labour demand, generating an increase in output, but not necessarily in the real wage, given demand determined unemployment. The increase in real wages occurs if the rise in public spending leads to a 'productivity enhancing' effect or through assumptions of complementarity between private and government consumption, rule-of thumb-consumers and sticky prices[2] (Gali et al., 2007). The employment consequences of fiscal policy are affected by labour market characteristics (Algan et al., 2002; Forni&Giordano, 2003; Ardagna,2007; Gomes, 2009; Monacelli et al, 2010; Holden&Sparman, 2011). The 'wage

[1] New classical suggest that labour productivity declines since extra public spending increase the interest rate , thereby crowding out private investments and reducing the capital stock.

[2] Rule-of-thumb consumers partly insulate aggregate demand from the negative wealth effects generated by the higher levels of taxes needed to finance the fiscal expansion, while making it more sensitive to current disposable income. Sticky prices make it possible for real wages to increase or, at least, to decline by a smaller amount even in the face of a drop in the marginal product of labour.

pressure' effect arising from a fiscal stimulus may be stronger in a non-perfect labour market where unions operate. For instance, increases in public employment increase the probability of workers being employed in the public sector rather than the private sector, so unions ask for higher wages in the private sector. The increase in labour costs decreases employment in the private sector and the rate of return on capital, leading to a decrease in output.

Recent theoretical model based studies try to analyze the effects of fiscal policy on output controlling for labor market institution characteristics and one study (Monacelli et al (2010) does this in a VAR context. The evidence is mixed. Algan et al, (2002) and Forni & Giordano (2003) examine the effect of public employment in unionized labour markets using partial equilibrium models. The former study suggests that an increase in public sector employment can reduce unemployment, if public sector wages are low, while the later suggests that the crowding out effect on private employment due to a public employment increase appears to be stronger in the presence of encompassing union behaviour. Ardagna (2007) study the issue in a dynamic general equilibrium model with a unionised labour market, suggesting that an increase in public sector employment, wages or unemployment benefits, raises the wage in the private sector and thus unemployment. Holmlund and Linden (1993), exploring the effects of public employment in a calibrated search model, find ambiguous effects on unemployment, since the positive direct effect of reducing the level of unemployment may be offset by the increased wage pressure due to increased public employment. Monacelli et al. (2010), exploring the effect of government consumption in a neoclassical model augmented with search and matching frictions, suggests that while higher government consumption increases the hiring rate due to the negative wealth effect inducing higher labor supply, this effect is crowded out by the rise in the real interest rate. Pappa (2009) finds that real wages and employment increase in response to government consumption and investment shocks. However, the effect of public employment is ambiguous. Linnemann (2009) suggests that increased public employment leads to increased total employment, while Bruckner and Pappa (2010), in an analysis of 10 OECD countries using structural VARs, find that higher government expenditures increase the unemployment rate. Gomes (2010) finds mixed effects of fiscal shocks on unemployment in a DSGE model with search and matching frictions. Holden&Sparman (2011) investigate the effects of fiscal policy in a regression framework using a panel dataset, controlling for other explanatory variables[3], like the cyclical situation of the economy, the openness of the economy, the type of fiscal impulse, monetary regime and labor market institutions. They find that an increase in government purchases leads to a significant reduction in unemployment, while the effect is higher for economies in recession and those using a fixed exchange rate.Studies reporting employment multipliers are few. Hence, this MRA analysis is focused on output fiscal multipliers.

2.1.2 Empirical evidence

The effectiveness of fiscal policy on stabilizing the economy has been investigated in the literature by using theoretical based models (RBC, NK-DSGE and MACRO) as well as econometrically based models (SEE and VAR). A clear distinction should be made between calibrated and estimated models, because they imply different approaches for deriving fiscal multipliers, and may make the findings of these categories not comparable to each other. RBC and NK-DSGE are structural micro-founded DSGE models based on the assumption of rational and forward looking agents. The model is solved if the resulting optimisation

conditions together with the equilibrium conditions are linearized around the steady state, and, the dynamics of the variables following a fiscal shock, are investigated by calibrating the model parameters to values based on microeconometric evidence or theoretical justifications. Large scale macroeconometric models (MACRO) are built from many equations which describe relationships derived from empirical data. They are like structural models in that they also use economic theory to limit the complexity of the equations. A quantitative assessment of fiscal policy shock is derived by simulation.

Many empirical studies use various single equation estimations (SEE hereinafter) such as OLS, IV, ML, GMM, and ECM approaches and report a wider range of multipliers. The main regression of these studies is in the following stylized form:

$$\Delta Y_t = \beta_0 + \beta_1 \Delta G_t + \sum X_t + \varepsilon_t$$

where, Y_t is per capita real GDP, G_t presents the fiscal shock and X_t is a set of additional control variables that may affect the relationship between output and the fiscal impulse. Due to simultaneity problems, usually the lagged values of G_t or IV for G_t are used in the above regression. The multiplier in single equation estimations usually appears in the coefficient, β_1, on the (lagged) fiscal variable.

Another strand of the literature applies VAR models using various methods of identification of exogenous fiscal shocks. A structural VAR model is in the following form:

$$AX_t = B(L)X_{t-1} + \varepsilon_t$$

where, X_t includes the endogenous variables of interest, A is the matrix of coefficients capturing the contemporaneous relationships, B(L) captures the dynamic relationships between endogenous variables and ε_t is a vector of orthogonal structural shocks.
A reduced form VAR takes the following form:

$$X_t = C(L)X_{t-1} + u_t$$

where, C(L) is a combination of A and B(L) matrixes and u_t is a vector of reduced form residuals. The relationship between the structural and reduced form residuals is given by:

$$\varepsilon_t = Au_t$$

Fiscal policy is potentially endogenous, since its key indicators depend on the state of the economy. The main challenge while studying the effectiveness of fiscal policy is to identify the exogenous discretionary policy by imposing restrictions on A, and in the literature this issue is tackled by using one of the four main approaches:

1) **The narrative approach:** exogenous fiscal shocks are identified using different sources of information in order to ensure that shocks are unrelated to macroeconomic developments and not anticipated. Ramey and Shapiro (1998) and Ramey (2011) use news reports in BusinessWeek and other sources to identify military buildups and other changes in US government purchases. Romer and Romer (2010) uses real time information such as government announcements or economic forecasts to identify the key motivation of each postwar legislated tax change in the US. IMF (2010) identifies fiscal policy actions in fifteen OECD countries implemented to reduce the budget deficit, and thus unrelated to economic activity in the short run.

2) **The recursive VAR approach**: identifies fiscal shocks by Choleski decomposition which imposes zero restrictions in order to implement a causal order of the VAR variables and to avoid contemporaneous reactions of the fiscal variable to business cycle variations (Fatás and Mihov, 2001).
3) **The Blanchard and Perotti (2002) SVAR approach**: identifies fiscal shocks by using institutional information and the assumption of decision lags. The difference of this approach from the recursive VAR approach is in allowing for non-zero restrictions such as estimated elasticities of automatic stabilizers.
4) **The sign restricted approach**: fiscal shocks are identified by imposing sign restrictions to the impulse-response functions of the fiscal shocks for a few periods in order to be distinguished from a business cycle shock. For instance, Mountford and Uhlig (2009) uses the following restriction: the impact responses of the fiscal variable to the fiscal shock and of the output variable to the output shock to be non-negative; the response of the fiscal variable to output is restricted to be non-negative; and the contemporaneous reaction of the output shock to the fiscal shock is left unrestricted.

After identifying the shocks, endogenous variable can be rewritten into the moving average form, and the dynamic responses to the structural shocks can be obtained. These are usually presented in the form of impulse response functions of standardized fiscal policy shocks.

The empirical literature on the size of the fiscal multipliers, in the post crisis period, is growing fast. Studies use several model classes, identification strategies and different specifications. However, the results are far from consensus. The heterogeneity of reported fiscal multipliers might be attributed to structural characteristics of different economies and to the settings and method of analysis. The size of the multiplier may be affected by the monetary policy reaction, openness of the economy, the indebtedness of the country, financial markets developments and labor market characteristics of the country. In terms of fiscal shocks, the effectiveness of fiscal stimulus depends on its composition. The empirical results are mixed. Some studies find higher multipliers for tax cuts or transfers, while others suggest that government spending is the most effective policy on increasing output. (Gechert & Will, 2012)

In the following, a brief summary of empirical findings and a selective discussion of studies presenting the main strands in VAR literature will be provided. Empirical studies find both Keynesian and neoclassical effects. The divergence of multipliers may arise due to the type of impulse, different data base or identification approach used (Caldara & Kamps, 2008; Favero & Giavazzi, 2012). Empirical evidence suggest that despite of the chosen identification approach several VAR studies agree that positive government spending shocks have positive effect on output. Ramey and Shapiro (1998) use a narrative approach to identify shocks to military spending and report an increase in output, but a decline in consumption. Fatás and Mihov (2001) and Blanchard and Perotti (2002) suggest that a fiscal expansion generates a positive response in output, with the associated multiplier being greater than one in the former, but close to one in the latter. Both studies suggest a large and significant increase in consumption. Using the sign restriction approach, Pappa (2009) and Mountford and Uhlig (2009) suggest a positive effect of government spending on output. Pappa (2009b) reports also a positive effect on consumption in contrast to Mountford and Uhlig (2009) who find that the response of consumption is zero and insignificant. Ramey (2011) using the narrative approach also suggests a negative effect on consumption. A similar effect is reported by Tenhofen and Wollf (2010). Regarding tax shocks, studies using sign restriction (Mountford and Uhlig, 2009) or a narrative approach to isolate tax changes unrelated to the

state of the economy (Romer and Romer, 2007) conclude that unanticipated tax increases have strong negative effects on output and consumption. In contrast, studies using the structural VAR approach yield mixed results. Blanchard and Perotti (2002) finds a negative effect of tax shocks on output whereas, Perotti (2005) results suggest that output does not react in the U.S. in the period when the tax shock hits the economy.

The effects of fiscal policy seem to change over time. Blanchard and Perotti (2002) find that estimated multipliers become unstable if different decades are dropped in a sample of US data spanning from 1947 to 1997. Perotti (2005) studies two separate sub-samples of five OECD countries suggests that the effects of fiscal shocks on output and consumption have decreased during the 1980s relative to previous decades.

A critique addressed to the VAR model findings is that due to implementation lags fiscal shocks may be anticipated, hence VAR results may be biased. Ramey (2008) suggests that in cases failing to account for the anticipation effect, the SVAR model captures shocks too late i.e. the initial decline in consumption that occurs as the news is received is missed. Blanchard and Perotti (2002) and Heppke-Falk, Tenhofen and Wolff (2006) include an indicator of future fiscal policy measures and suggest that findings are qualitatively similar with VAR studies not accounting for anticipation effects.

Recent VAR studies augment the conventional VAR model with variables to control for the effect of government debt dynamics (Galí et al, 2007; Ilzetzki et al, 2012; Favero and Giavazzi, 2007), the degree of openness (Favero et al, 2011; Crespo et al, 2011; Brückner and Tuladhar, 2010), and labour market institution characteristics (Monacelli et al, 2010). An important strand in VAR studies are non -linear VAR models used in order to distinguish the effect of fiscal policy according to the state of economy (Baum and Koester, 2011; Auerbach and Gorodnichenko, 2012; Burriel et al, 2010; Cimadomo and Bénassy-Quéré, 2012; Hernández de Cos and Moral-Benito, 2013; Baum et al, 2012; Corsetti et al, 2012; Bachmann and Sims, 2012). Debt dynamics is an important factor determining the effectiveness of fiscal policy. The omission of public debt in the VAR leads to biased results as they fail to take into account the debt dynamics that arise after a fiscal shock and exclude the possibility of taxes and spending responding to the level of debt （Favero and Giavazzi, 2007; Corsetti, Meier, and Mueller, 2009; Leeper, 2010; and Zubairy, 2010). Favero and Giavazzi (2009), include government debt changes in the model and suggest that there are not significant differences in estimates with and without the inclusion of a non-linear government budget constraint. Ilzetzki et al (2012) investigates the differences in fiscal multipliers arising from different country characteristics, by doing subsample analysis on 44 countries, using a conventional VAR model. Their findings suggest higher multipliers in economies with fixed exchange rates and open economies and negative multipliers in high-debt countries. Corsetti et al (2012),using a sample of 17 OECD countries for the period 1975-2008, reports similar results: higher fiscal multipliers in fixed exchange rate regime countries, facing a financial crisis, or under sound public finances. Auerbach and Gorodnichenko (2012) use a regime switching VAR model to analyse whether multipliers differ in recession and expansion. They find that, in a model allowing an endogenous switch between regimes, multipliers are typically lower than one during expansions, however, higher than one in recessions. A similar methodology was used by Baum and Koester (2001), Batini et al. (2012) and Baum et al. (2012) and these studies suggest larger multipliers during recession.

Several survey studies provide a descriptive approach to reported multipliers and characteristics of the reporting studies in the literature (Ramey 2011; Parker 2011; Hebous 2011; Spilimbergo et al. 2009). Rusnak (2011) conducts a MRA on studies using a VAR framework, augmented with primary data on structural characteristics of the economy,

suggesting that the reported spending multipliers systematically depend on the characteristics of the economy, and, to a lesser extent, on the method of analysis. Gechert & Will (2012) apply MRA to a set of 89 studies using various model classes, suggesting that the reported size of the fiscal multiplier crucially depends on the setting and method chosen. A novel approach of this study will be to augment MRA, despite the primary data on structural characteristics, with primary data on labour market variables to see if there is any systematic variation to be attributed to labour market institutions. The theoretical rationale for including each primary data variable is provided in the following section.

2.2 The dataset and MRA variables

The first step of conducting this MRA analysis on fiscal multipliers is identifying all set of published or unpublished studies that investigate the effect on fiscal policy on output and provide calculations of multiplier effects. Following Stanley and Doucouliagos (2012) recommendation the EconLit database and working paper series (NBER, IMF, ECB, CEPR) were searched using 'fiscal multipliers' and 'fiscal policy' as key words. Additionally, the references of already identified studies were investigated as well as studies citing the seminal paper of Blanchard and Perotti (2002). We have identified over 101 empirical and estimated/calibrated papers on short-term output effects of discretionary fiscal policy measures, for the period 1992-2013 and collected 1363 observations of output multipliers. We have counted 50 observations from RBC models, 86 observations from NK-DSGE models, 313 observations from MACRO models, 855 observations from VAR models and 59 observations from SEE.

The MRA dataset includes only empirical studies (VAR and SEE) and consists of a total of 914 observations. The reported multipliers from estimated/calibrated papers (RBC, NK-DSGE and MACRO) are not included in the main MRA analysis. The purpose of gathering these estimates is for comparison with the MRA. The following section provides descriptive statistics of these studies. The rationale for excluding these studies is that these models are not estimated empirically and are not for a particular economy/country. RBC, NK-DSGE and MACRO do not provide any measure of precision and not provide information on sample size, which are necessary to apply the publication bias test. Additionally, classifying the estimates into RBC or NK-DSGE models is not always straightforward and implies a considerable degree of subjectivity by the meta-analyst.

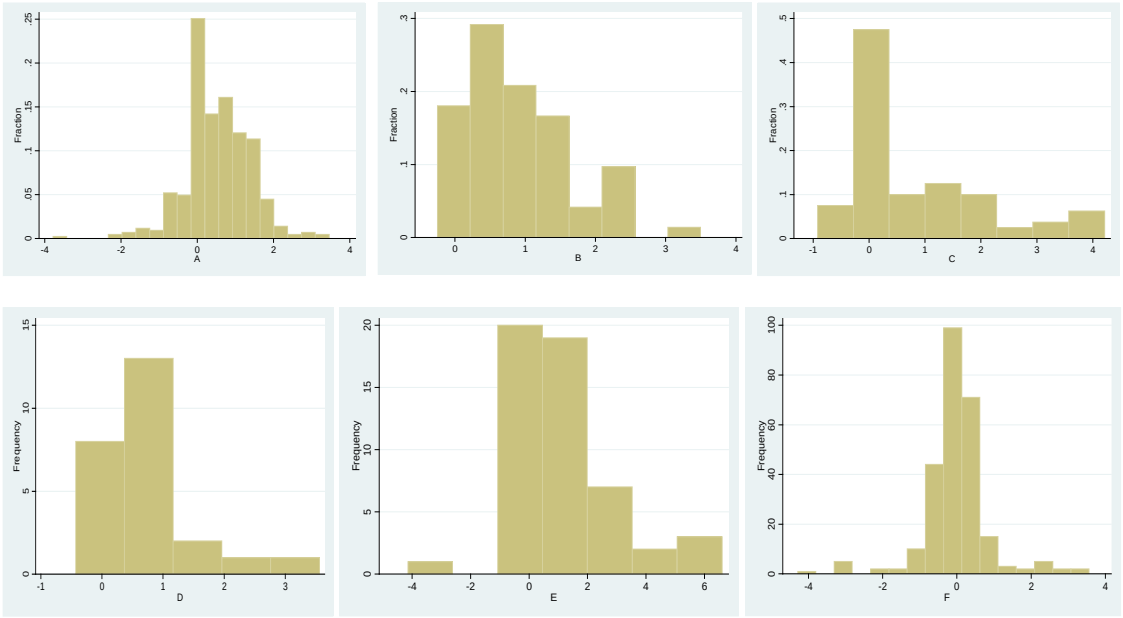
2.2.1 Descriptive statistics

This section provides descriptive statistics on reported multiplier values for all identified studies. The mean value of reported multipliers for the total sample is 0.6. The null hypothesis that the mean multiplier is equal to zero is rejected (t-stat=19.44). Table 1 reports mean values of reported multipliers with respect to model classes, fiscal impulses and various identification strategies of fiscal shocks, suggesting that multiplier values vary widely among these categories. MACRO models seem to report the highest multipliers, while those from SEE seem to be the lowest. Means are in a range of 0.31 to 1. Fiscal impulses also differ in terms of reported multipliers. Public employment and public investments seems to be the most effective fiscal shocks reporting mean multipliers above 1 while taxes seems to be the least effective reporting mean multipliers close to 0. Means are in a range of 0.01 to 1.13. Regarding different identification strategies of exogenous fiscal shocks in VAR models, studies using the recursive approach report higher multipliers, while those from the Blanchard and Perotti (2002) SVAR approach report the lowest. Means are in a range of 0.41 to 0.89. Figures 1-3 suggest that multipliers for the above categories are not normally

distributed. However, the simple average might present a distorted picture about the true value of multipliers. The underlying size of the multiplier might differ across countries due to structural characteristics or the difference may arise due to method of investigation of studies. A formal MRA is needed in order to separate structural from method-specific effects. The considerable variation around the mean, from negative to positive values supports the use of MRA to explain the heterogeneity of reported multipliers.

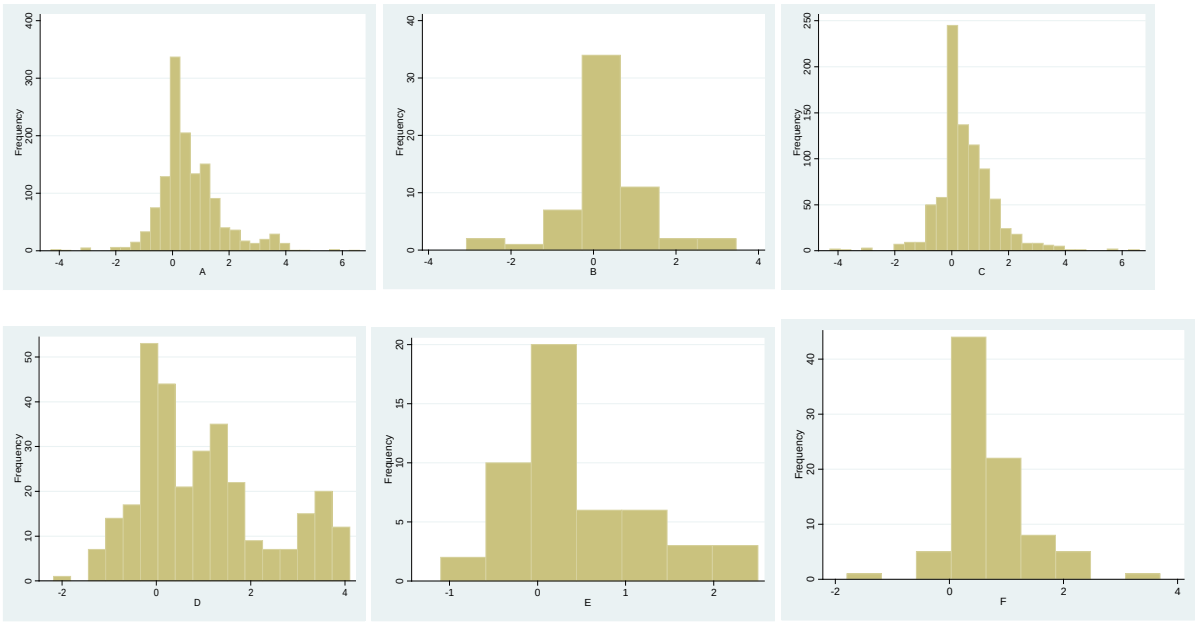
Table 1: Descriptive statistics of reported multiplier values for model classes, fiscal impulses and identification strategies

Fiscal Impulse	**N**	**Mean**	**Std. Dev**	**Min**	**Max**
Government spending (not specified)	423	0.53	0.85	-3.78	3.46
-Consumption	72	0.91	0.74	-0.25	3.5
Investment	80	1.08	1.2	-0.92	4.21
Military spending	25	0.71	0.86	-0.43	3.56
Public Employment	52	1.13	1.77	-4.14	6.6
Taxes	263	0.01	0.89	-4.29	3.57
Model class	**N**	**Mean**	**Std. Dev**	**Min**	**Max**
Total sample	1363	0.61	1.11	-4.29	6.6
SEE	59	0.31	1.05	-3.08	3.46
VAR models	855	0.50	1.01	-4.29	6.6
MACRO models	313	1	1.38	-2.18	4.1
RBC models	50	0.39	0.76	-1.1	2.5
NK- DSGE models	86	0.67	0.75	-1.8	3.7
Identification strategy	**N**	**Mean**	**Std. Dev**	**Min**	**Max**
SVAR	620	0.41	0.86	-2	4.21
Narrative	34	0.44	1.82	-4.29	3.5
Recursive	91	0.89	0.56	-1.97	3.34
Sign restrictions	110	0.69	1.43	-4.14	6.6



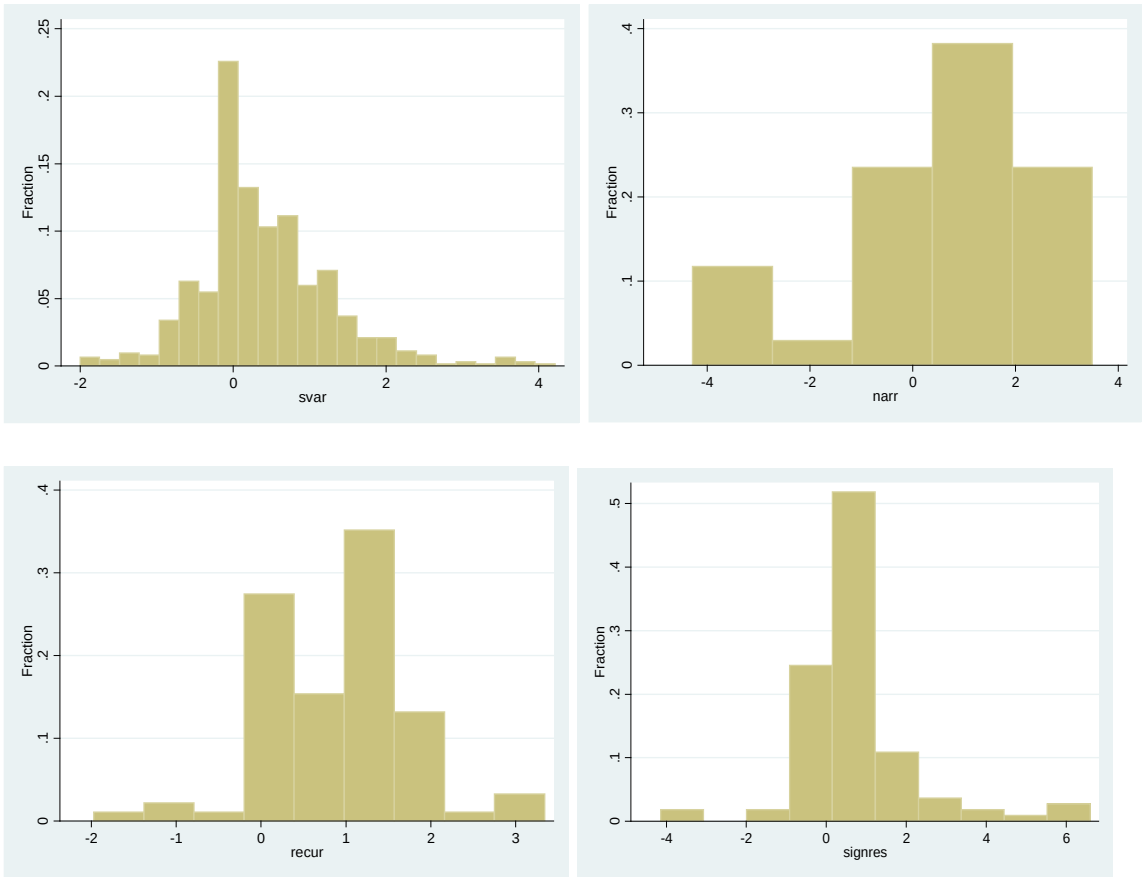
A- Government spending (not specified); B- Consumption; C- Investment; D-Military spending; E- Public Employment; F-Taxes

Figure 1: Histograms of reported multiplier values for various fiscal impulses



A- Total sample; B- SEE; C- VAR models D-MACRO models; E- RBC models; F- NK-DSGE models

Figure 2: Histograms of reported multiplier values for the total sample (1334 observations) and various model classes



svar- B&P (2002) SVAR approach; narr- Narrative approach; recur-Recursive approach; signres- Sign restriction approach

Figure 3: Histograms of reported multiplier values for various identification strategies

2.2.2 Description of MRA variables

In order to explain the differences in reported multipliers we develop a set of moderator variables that capture the study characteristics and add primary data for structural characteristics of the economy. The choice of moderator variables is based on the typical characteristics discussed in the theoretical literature. A description of variables, as well as the rationale, for using these moderator variables is provided below.

Study characteristics moderator variables

For the total sample we use the following variables:

1) **Type of model class**- dummy variable;
   VAR-1 for estimates from VAR models, 0 otherwise;
   SEE-1 for estimates from SEE models, 0 otherwise;

   Rationale: Coding for the type of model class is more sensible in the case of theoretical based model, since they rely on different assumption, which are crucial for the effect of fiscal policy. Estimates in our sample are mostly from atheoretical VAR models, however few studies are estimated by a single equation (in most cases dynamic panel estimated by GMM is used). Hence, we code for the model class to see whether there is a significant difference between VAR and SEE multipliers.

2) **Type of fiscal impulse**- dummy variable
   CONS- 1 if the impulse is government consumption, 0 otherwise;
   INVEST- 1 if the impulse is government investment, 0 otherwise;
   MILITA- 1 if the impulse is military spending, 0 otherwise;
   PUBEMP- 1 if the impulse is public employment, 0 otherwise;
   TAX- 1 if the impulse is the change in taxes, 0 otherwise;
   NOTSPEC- 1 if the impulse in the primary study is not specified (general government spending), 0 otherwise;

   Rationale: theoretical predictions suggest that the effectiveness of fiscal policy is dependent on the type of fiscal impulse used. For instance, Keynesian theory suggests that tax multiplier is smaller than the government spending multiplier, because, due to marginal propensity to save, a part of the increase in disposable income will be saved and not directly spent.

3) **Direction of the impulse**- dummy variable
   INCR- 1 if the impulse shock is positive, 0 otherwise;

   Rationale: Studies in our sample investigate the effects of positive fiscal impulse shocks as well as negative ones. Coding for the direction of the impulse is crucial

since the effects are not supposed to be symmetric, but, they are linked/ dependent with the state of the economy. For instance, in a recession economy, a tax increase will not have a symmetric effect as a tax cut.

4) **The way fiscal shocks are financed-** dummy variable
DEBTFIN-1 if the shock is debt financed, 0 otherwise;

Rationale: the effects of the shocks financed by distortionary taxation are significantly different from the effects of the shock financed by lump-sum taxes. If fiscal shocks are entirely financed by distortionary taxation, fiscal policy is ineffective and may even result in lower output since the lower after-tax factor reward may affect the level of investment in the economy (Baxter and King, 1993).

5) **The duration of the shock**-dummy variable
TEMPOR- 1 if the duration of the shock is temporary, 0 otherwise;

Rationale: Baxter and King,(1993) suggest that temporary versus permanent shocks tend to have different effects i.e the effect is higher if the shock is permanent.

6) **Type of the country**-dummy variable
TRANSIT-1 if the effect is estimated for transition countries, 0 otherwise;

Rationale: few studies in out sample have estimated fiscal multipliers for transition economies: Czech, Hungary, Poland, Slovakia, Slovenia, Croatia and Bulgaria. Fiscal multiplier is expected to be lower in these countries since they are small open economies. The rational is that in small open economies with less developed financial markets, the sovereign risk premium is higher which will result in fiscal shock having a stronger effect on interest rates, and consequently, offsetting the initial impulse (Muir and Weber, 2013).

7) **Type of data**-dummy
QUART-1 if quarterly data used, 0 otherwise;
ANNUA-1 if annual data used, 0 otherwise;

Rationale: the frequency of the data is important for VAR studies in terms of identifying exogenous and unanticipated fiscal shocks (Blanchard and Perotti(2002); Ramey (2008)). Hence, we code for the properties of time-series to see whether there is a significant difference between multipliers estimated by quarterly and annual data.

8) **Horizon of estimation**- continuous variable
HORIZ- number of quarters after the shock on which the multiplier calculation is based

Rationale: due to implementation lags of fiscal policy, it is expected that the longer is the horizon of estimation, the higher is the value of the multiplier.

9) **Type of fiscal multiplier**-dummy variable
CUM-1 if the effect is a cumulative multiplier, 0 otherwise;
PEAK-1 if the effect is a peak multiplier, 0 otherwise;
IMPA-1 if the effect is an impact multiplier, 0 otherwise;

Rationale: multipliers in our sample are calculated either as a peak response of GDP to the initial shock g; a cumulative response of GDP_{t+n} to the cumulative shock g_{t+n}; or as an impact response of GDP to the impact impulse. We code to see whether there is a significant variation of multipliers attributed to the method of calculation.

10) **Controlling for country specific characteristics**- dummy variable
CONVEN-1 if the study uses conventional VAR variables[4], 0 otherwise;
INDEBT-1 if the study controls for the level of debt/GDP, 0 otherwise;
OPEN-1 if the study controls for the degree of openness, 0 otherwise;
ER-1 if the study controls for the exchange rate regime, 0 otherwise;
EMPLOY-1 if the study controls for the level of employment/hours, 0 otherwise;
LMI-1 if the study controls for the labor market institutions characteristics, 0 otherwise;
STATECO-1 if the study controls for the state of the economy (recession/expansion), 0 otherwise;

Rationale: as already discussed in the literature review, structural characteristics of the economy are important determinants of the effectiveness of fiscal policies. Recent studies in our sample control for these characteristics. We distinguish between these studies and conventional studies by using the above coding in order to investigate the differences in the reported multipliers.

11) **Controlling for the quality of the study**- dummy variable
SAMSIZ- continuous variable, number of observations;
SIGNIF-1 if the effect is significant, 0 otherwise;
PUBLISH-1 if the study is published in a peer- reviewed journal, 0 otherwise;
ROBUST-1 if robustness checks are done, 0 otherwise;
STATIO-1 if the study tests for stationarity, 0 otherwise;
DIAGN-1 if the study presents model diagnostics, 0 otherwise;

Rationale: Philips and Goss (1995) critics MRA that mixes results from poorly designed studies with those from well-designed studies. We control for the quality of the study by using the above coding.

Structural characteristics moderator variables

For the total sample we use the following primary data[5]:

[4] Most of the studies include in VAR analysis only three variables: government expenditure, government revenue and GDP.
[5] See Appendix A.19 for detailed primary data description and sources.

1) **Labour market institutions characteristics**:
   EPL-(index scaled 0-5), Employment protection legislation indicator measures the strictness of employment protection for the employer;
   TUD-(ratio), Trade union density is defined as the ratio of wage and salary earners that are trade union members, divided by the total number of wage and salary earners;
   BRR-(ratio), benefit replacement rates is a measure of how much unemployed worker receives in benefit from the government;
   COOR-(index scaled 1 to 5)- the coordination of wage bargaining is based on Kenworthy's 5-point classification of wage-setting coordination scores;
2) **The indebtedness of the economy:**
   DEBTGDP-(%), central government debt/ GDP
3) **Monetary policy reaction**:
   IR-(%), short term money market rates
4) **The degree of openness**:
   IMPGDP-(%), imports of goods and services/GDP
5) **Financial development**:
   CREDGDP-(%), domestic credit to private sector/ GDP;

   Rationale: the arguments why labour market characteristics and other structural variables, are important for fiscal policy effectiveness are already provided in the literature review.

For the VAR subsample we additionally control for the following variables:

1) **Identification strategy**:-dummy variable
   NARR- 1 if the narrative identification is used, 0 otherwise;
   RECUR-1 if the recursive identification is used, 0 otherwise;
   SVAR-1 if the Blanchard and Perotti (2002) SVAR identification is used, 0 otherwise;
   SIGNRES-1 if the sign restriction identification is used, 0 otherwise;
2) **Type of VAR model**: dummy variable
   TIMESER-1 if time series VAR model is used, 0 otherwise;
   PANEL-1 if panel VAR model is used, 0 otherwise;
   NONLIN- 1 if non-linear VAR model is used, 0 otherwise;

2.3 MRA methodology

Meta-analysis is a quantitative analysis of estimated effects of interest in an empirical literature which aims to summarize findings from the empirical literature and to identify sources of heterogeneity in these findings and also identify and control for publication selection bias in the empirical literature (Pugh et al, 2011). Thus, the meta-analysis uses results from primary studies as the data for analysis. MRA is a rigorous, objective and clarifying approach to identifying trends or representative values in the investigated literature as well as to identifying what determines different values and can be used as a complement to

narrative literature review. MRA is particularly used in literatures that report a wide range of results and where is heterogeneity in the institutional settings and econometric models adopted. Since primary studies usually are specified in different ways and use different measures of variables, the first challenge for the meta-analyst is to identify effect sizes that are comparable across different studies. Meta studies often use standardization tools to construct the effect size. In our case standardization is not necessary, because the multiplier is already dimensionless. Further, in order to account for the differences between impact, peak and cumulative multipliers we control for the multiplier calculation method and the time horizon to extract comparable multiplier values.

In order to explain the heterogeneity in estimates in our sample of studies, we begin with the following meta-regression model suggested by Stanley and Jarrel (1989):

$$k_j = k_0 + \sum_{k=1}^{k} \alpha_k Z_{jk} + e_j \qquad j = 1,2,\ldots L, \text{ where:} \qquad \text{Eq (1)}$$

- k_j is the multiplier value of observation j;
- k_0 is the "underlying" or "reference" multiplier value;
- Z_{jk} is the vector of characteristics ("moderator variables") of observation j;
- α_k is the vector of systematic effects of Z_{jk} on k_j;
- e_j is the meta-regression disturbance term;
- L is the number of studies.

In terms of Z_{jk} we have decided to include primary data. Our innovation is to apply MRA in order to explore dimensions that were not considered by the primary studies. We take advantage of the between-study heterogeneity to provide further insights into the effects of fiscal policy and to explore additional dimensions. We do this by collecting information on structural characteristics of the countries investigated by primary studies. Hence, we assess the evidence base by drawing upon data from within the studies themselves- reported multiplier estimates, as well as information that was (partially) not considered by the authors, such monetary policy reaction, openness of the economy, the indebtedness of the country, financial markets developments and labor market characteristics of the country, at the time of the analysis. This allows us to model the heterogeneity within the primary studies, through their study specific characteristics, as well as the heterogeneity in the samples used by different studies which was not previously modelled by the studies.

Primary studies usually provide multiple estimates of output multipliers for various models, countries or types of fiscal impulse. Multiple estimates reported within a single study might not be statistically independent of each other, violating one of the OLS assumptions. Additionally, different estimates used from a study may cause an undue weight of a single study. Stanley (2001) suggests to use only one observation per study or to take the average in order to tackle these problems. However, following this strategy has its own drawbacks: it imposes a trade-off with variability and degrees of freedom, subjectivity by the meta-analyst which value to take, undue weight of more comprehensive studies (Bijmolt and Pieters (2001)) Following other authors (De Grauwe and Costa Storti(2004); Nijkamp and Poot (2004); Card et al. (2010); Rusnák et al. (2011); Gechert & Will (2012)) we have decided to use multiple estimates per study in this analysis. Therefore, we adjust the standard errors for data clustering, using each study in our dataset as a distinct cluster. Since the studies in the primary literature may use different data sets, different sample sizes, and different independent variables, the variances of the estimated regression coefficients will not be equal, leading to possible heteroscedasticity in MRA (Stanley and Jarrel, 1989).In order to tackle the problem of heteroscedasticity Stanley and Jarrel, (1989) suggest to divide the Eq(1) by

the standard errors of k_j. However, due to lack of data for standard errors in our case, it is not possible to follow the mainstream approach in this analysis. In order to tackle the problem of undue weight of the studies reporting the largest number of estimates, each estimate is weighted by the inverse of number of estimates in a given study.

2.3.1 Publication selection bias

Publication selection bias occurs when authors do not report all of the results they uncover; rather, they selectively report statistically significant results or results consistent with a certain theory, which they believe have a stronger chance of being published. Consequently, the inference arising from the empirical evidence may be biased.

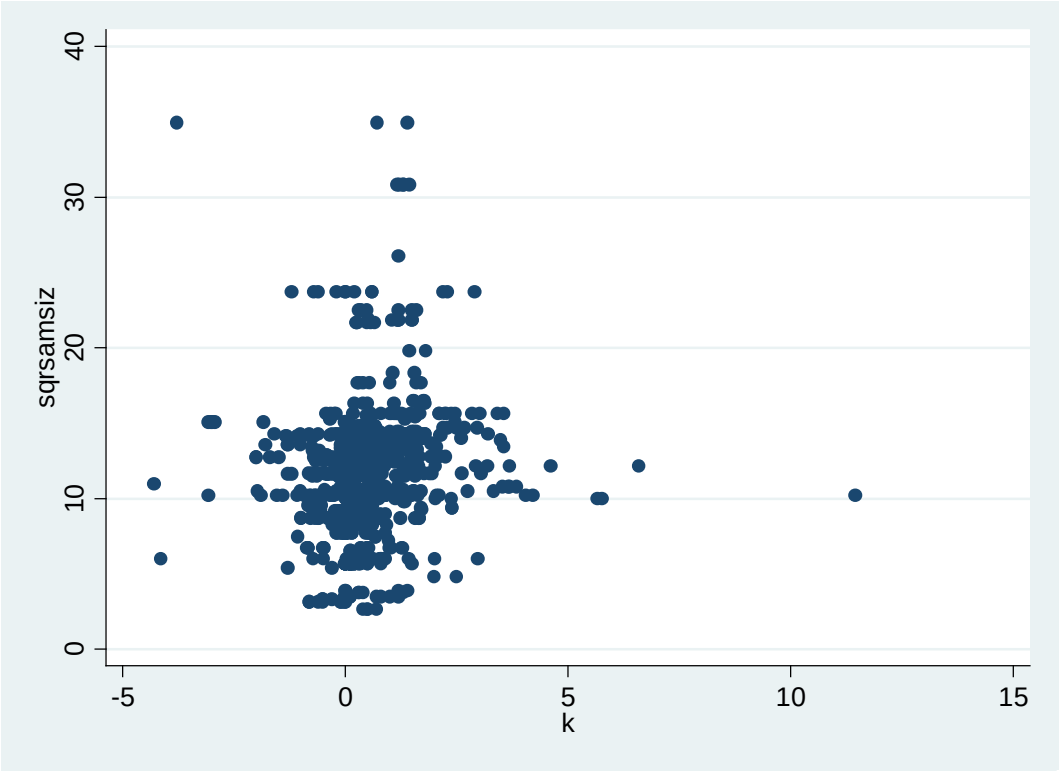


Figure 3: Funnel plot test

In order to investigate for publication selection bias in the fiscal multiplier literature, as suggested by Stanley & Doutciloagios (2012), we start by analysing the funnel plot in Figure.3.The estimated multiplier values from the literature are placed in the horizontal axes while in the vertical axes is placed the square root of sample size which as sampling theory suggests is a proxy for precision of the estimates. In an empirical literature free of publication bias the plot will resemble an inverted funnel; namely, symmetric, widely spread at the base (reflecting a lack of precision in small sample studies) but to a narrow range of estimates around the mean of the sample (reflecting more precise results from a larger sample studies).In this case, the distribution looks more or less symmetric, slightly skewed to the right, around a mean that is positive but small, hence, suggesting that there is a mild case or even no publication bias in this literature. However, despite the funnel plot analysis, a formal test of publication selection bias is needed. Stanley (2005, 2008) suggests a simple and

powerful, test for publication bias – the funnel asymmetry precision effect size test (FAT-PET) which involves regressing the effect size against a constant and the effect size's standard error (*SEij*).In case of no publication selection bias in the literature, the estimates should not be correlated with their standard errors. On the other hand, if researchers search for estimates that are statistically significant, then they will re-estimate their models until the relationship between the estimates and *SE* achieves some acceptable standard of statistical significance. This process will generate a correlation between the effect size and their standard errors (Stanley, 2008). Unfortunately, as reported above, most studies of fiscal multipliers do not report either the standard errors or *t*-values on the estimated effects. Consequently, it is not possible to investigate publication bias in the fiscal multipliers literature with these standard methods. Following Velickovski and Pugh (2011) and the suggestions of Stanley and Doucouliagos (2012) that the square root of the sample size can serve as a rough proxy for precision in FAT-PET-MRA when the standard errors are unavailable, we explore the possibility of publication selection bias in this literature by estimating the following regression:

$$k_j = k_0 + \beta_1\left(\frac{1}{\sqrt{N}}\right)_j + e_j \qquad \text{Eq (2)}$$

where, k_j is the multiplier value of observation j, $\left(\frac{1}{\sqrt{N}}\right)_j$ is the inverse of the square root of the sample size (N) in regression j and e_j is the error term.[6] In Eq (2), k_0 estimates the true effect. In the above regression we test whether there is a relationship between the magnitude of the effect and the sample size. If the null hypothesis that $\beta_1=0$ is not rejected, then there is no evidence that larger sample estimates are systematically different in magnitude from smaller sample estimates. This suggests an absence of publication bias in the literature. Sampling theory suggests that, as the sample size (N) increases, estimates of k_j approaches or converges to their true value k_0 (Gujarati, 2004). Larger-sample estimates should be more precise than smaller-sample estimates. Using the above principle that larger-sample estimates are more precise but not necessarily higher in magnitude than smaller-sample estimates as well as the fact that publication bias arises if the researcher repeats imprecise estimation as long as it gets effects higher in magnitude in order to get significant estimates, we expect to find the trace of such selection procedures in systematic variation of effect size with sample size. The results suggest that the null hypothesis that $\beta_1=0$ is rejected. In our model $\beta_1=-2.24$, (p-value=0.03) is significantly different from zero, suggesting that there is a systematic variation of effect size with sample size.

However, in order to avoid the potential problem of endogeneity bias that may arise if $\left(\frac{1}{\sqrt{N}}\right)_j$ is correlated with the moderator variables $Z_{jk,}$ we include $\left(\frac{1}{\sqrt{N}}\right)_j$ in Eq (1) and estimate:

$$k_j = k_0 + \sum_{k=1}^{k} \alpha_k Z_{jk} + \beta_1\left(\frac{1}{\sqrt{N}}\right)_j + e_j \qquad \text{Eq (3)}$$

[6] We have decided to use $\left(\frac{1}{\sqrt{N}}\right)_j$ the inverse of the square root of the sample size (N) instead of $\sqrt{N}$ on the grouds that the model including $\left(\frac{1}{\sqrt{N}}\right)_j$ has more satisfactory model diagnostics compared to a model including $\sqrt{N}$.

The results in Table.3 replicate the findings from the bivariate regression estimated in Eq (2). β_1=-2.37, (p-value=0.01) is significantly different from zero, suggesting that there is a systematic variation of effect size with sample size. Hence, considering both results, we can conclude that the literature is prone to publication selection bias. The interpretation of the slope of a reciprocal model is counterintuitive ($dY/dX = -\beta_1(1/X^2)$, implying that if β_1 is positive the slope is negative throughout and if β_1 is negative the slope is positive throughout (Gujarati, 2007, p.188). In this case the estimate of β_1=-2.37 indicates positive publication bias (as suggested by the funnel plot).

2.4 MRA results

In this section, the results for the total sample of VAR and SEE estimates and for the subsample of VAR studies will be presented and discussed. Moderator variables measured on a nominal scale are multicollinear since each observation must belong to one value in this group, therefore, we proceed by omitting one variable of a closed group. The influence of these omitted variables is reflected in the constant, k_0. Following Stanley and Doucouliagos (2012) we apply a general-to-specific approach and firstly use all the moderator variables in the model. However the diagnostics are flawed and we exclude several insignificant moderator variables[7]; namely those controlling for the type of fiscal multiplier (cum, peak) and the way fiscal impulses are financed (debtfin).Also moderator variables controlling for the quality of the study are excluded (signif, publish, robust, statio, diagn); they are suggested by common practice in MRA, but most of them prove insignificant and are not variables of interest in this study. The results from the baseline model using Eq(3) are presented in Table.2.Standard diagnostic test from the ordinary least square suggest that the model is well specified with respect to linear functional form and normality, but may suffer from heteroscedasticity. Therefore, the model is estimated with cluster robust standard errors. Estimation of cluster-robust standard errors relaxes the independence assumption and requires only that the observations be independent across the clusters (Velickovski and Pugh, 2011).

The constant, k_0[8], of the baseline model in column (1) is an average multiplier value, calculated as a response to an unspecified public spending impulse, using a conventional SEE, not augmented with control variables for the characteristics of the economy, and estimated with annual data. The estimated value is higher than one and appears to be significant.

The results of the baseline model presented in column (1) suggest that model classes and fiscal impulses differ significantly concerning their influence on the multiplier. Estimated multipliers arising from the VAR model are significantly higher than estimates from SEE. Regarding the type of the impulse, public investment and public consumption produce higher multiplier values in our data set, while tax shocks have significantly lower impact compared

[7] See Appendix 1 for the results with all moderator variables included.

[8] k_0 presents an average multiplier for a study by all the omitted categories characteristics.

to unspecific/general public spending. For military spending and public employment the difference to unspecified public spending is insignificant. As expected, due to implementation lags in fiscal policy, a longer horizon of measurement comes with significantly higher multipliers. These results are consistent with the Gechert & Will (2012) findings. Studies using quarterly data report significantly lower multipliers compared to studies using annual data. The difference between a positive and a negative fiscal shock appears to be insignificant suggesting that there is no asymmetry in the effects due to the direction of the shock. The duration of the shock is highly significant suggesting that multipliers from a temporary shock are lower than multipliers from a permanent shock. The reported multipliers for transition countries do not differ significantly from multipliers for advanced economies. The results suggest that studies controlling for the degree of openness, type of exchange rate and labour market characteristics, yield significantly different estimates compared to conventional (no control variables for country characteristics) studies. For studies augmented with debt/GDP, employment and output gap variables, the differences in estimates compared to conventional studies is insignificant.

The results on structural characteristics are mixed. The degree of openness channel is highly significant, suggesting that due to the leakage effect of imports, more open countries have relatively lower multipliers that relatively closed economies. The effect of financial development, debt/GDP and interest rates appears to be insignificant, in contrast with a priori expectations. Regarding labour market institution characteristics, the benefit replacement rate has significant effects suggesting that in countries with higher replacement rates the multiplier effect is higher. The other labour market variables have an insignificant effect on estimated multipliers.

{ INSERT HERE Table 2: Total sample results}

Some robustness checks follow, by estimating the regression in column (1) in several variants. In the total sample there are two types of studies: 1) conventional studies which do not control for country structural characteristics; and 2) studies already controlling for country characteristics by augmenting the model with one/some structural variables. In order to investigate whether the insignificance of structural characteristic variables is due to this mix of heterogeneous studies, we estimate the model only for conventional studies estimates. The results are presented in column (2). However, in this specification all structural characteristics variables appear to be insignificant. The results for other variables are similar to the baseline model in column (1).Changing the reference group does not actually change much the results. Coefficients and significance levels are only altered very slightly in comparison to column (1).[9]

{ INSERT HERE Table 3: Subsample analysis: VAR results }

Table.3 presents estimates only from VAR studies. In column (1) the influences of various identification strategies and different approaches to the estimation of VAR models are investigated. However, standard diagnostic tests suggest that this model suffers from misspecification as well as from heteroscedasticity problems. In column (2) we re-estimate the model without control variables for the way of estimation of VAR models, since they

[9] As expected, using observations from VAR models as reference only affects the constant yielding a slightly higher value. The results are also robust to a different reference fiscal impulse. See Appendix A.4, A.5 and A.6 for the results with different reference groups.

prove insignificant in the first specification, and with cluster-robust SEs. The diagnostics improve; and the model appears to be well specified. The results in column (2) suggest that shocks identified via the recursive method generate significantly higher multipliers compared to SVAR identified shocks. For the shocks identified by the narrative and sign restriction strategies the difference of effects compared to SVAR identified shocks is insignificant. Compared to the total sample results (presented in Table 2), studies augmented with an employment variable turn out to report significantly higher multipliers than do conventional studies, while the rest of the results are similar.

A possible drawback of this model is not adding primary data in order to control for the state of the economy. Theoretically, multipliers should differ according to the state of the economy. Keynes suggests that fiscal policy is most effective in a depressed economy. Hence, to account for the state of the economy we would need to create a variable that captures the period/circumstances when the multiplier would be larger. However, we don't have data for depression period (or even a precise definition), hence it is hard to isolate those changes/shocks from normal business cycle data. Primary studies use output gap data to control for the state of the economy. However as we average data across the sample period, the average value of output gap will be not significantly different from zero, hence, not useful as primary data in our study. The nearest equivalent we can use to make an inference about the variation of effect size to the state of the economy, is to use the fact that Japan is included in our sample and consider its unusual economic development. Japan is an exceptional economy going from high level of growth to a long period of relative stagnation after the asset bubble crisis. Accordingly, Japan's economic development since 1989 can be considered as a proxy/approximation for a depressed economy. The data range of the relevant studies for Japan is the period 1970-2011. Hence, a dummy for all estimates from Japan is included in the main specification. Further, in order to control for the post financial crash period, another dummy, Japan90, is created for the estimates from Japan for the period after 1989 and added to the regression. The results, as shown in column (3), suggest that Japan90 is insignificant, while the first Japan dummy is significant (β=-0.939; p-value= 0.09). The negative coefficient of Japan dummy, suggesting a crowding out effect of fiscal policy, most likely reflects the pre-financial crisis period of high level of growth.

In this study the problem of not adding primary data for the state of the economy can be tackled, because the primary studies enable us to construct appropriate moderator variables to investigate the effects on estimated multipliers of different macroeconomic conditions. As explained above, recent studies controlling for the state of the economy, while investigating the effectiveness of fiscal policy, are accounted for by the moderator variable STATECO. However it should be noted that the number of this type of studies is small. In order to investigate more specifically the variation of multipliers with the state of the economy, the general variable STATECO is replaced with three new moderator variables:

RECC-1 if the multiplier is estimated assuming recession, 0 otherwise;
(7.4% of total observations)
EXP-1 if the multiplier is estimated assuming expansion, 0 otherwise;
(7.1% of total observations)
FINCRISIS- 1 if the multiplier is estimated assuming financial crisis, 0 otherwise;
(4.1% of total observations)

{ INSERT HERE Table 4: Total sample results- different specification}

The results of this different specification are presented in Table.4. The model in column (1) includes all three moderator variables and the results suggest that multipliers arising from studies assuming expansion are significantly different compared with estimates from studies not controlling for the state of the economy. The moderator variables for recession and financial crisis appear to be insignificant. However the diagnostics suggest that the model is miss-specified. Some robustness checks follow. The model in column (2) is estimated excluding the moderator variable for financial crisis, given that it proves insignificant and applies to very few observations. The diagnostics improve, while the results of EXP and RECC statistically do not change. In order to explore further the state of the economy we did some experiments, but the results don't change much and are presented in Appendixes[10]. Generally, in line with theoretical predictions, the multipliers estimated in expansion are smaller and the results are robust for different specification. The moderator variable for recession appears to be insignificant while the coefficient for financial crisis appears to be relatively big and positive but not stable, its significance varies whether the moderator variable for expansion is included in the model. Finally, the model in column (3) is estimated excluding the Japan dummies and the results are qualitatively the same. The moderator variable for transition countries turns out to be significant if the Japan dummies are included and vice versa. However we refrain from excluding the Japan dummies from the model, we consider they should be in the model controlling for Japan's exceptional economic development. Therefore, considering the above arguments as well as diagnostic results, our preferred model, for which we will continue further our analysis, is the model presented in column (2), Table.4. Compared to the first specification presented in column (3), Table.2, the coefficients of variables are qualitatively similar suggesting that the results are robust to different specifications.

The aim of this study is not to estimate the 'true' multiplier effect. The multiplier, as shown in this study, is time and state dependent. Even though a relatively large number of empirical studies are included in this analysis, given that theoretical studies are excluded, it is not clear whether this literature provides an unbiased representation of multipliers effects. However in terms of robustness checks and in order to investigate further the heterogeneity of reported multipliers, several representative effects for studies with different characteristics are provided in Table.5.The representative effects are calculated by setting moderator variables to zero, unless otherwise specified, and setting continuous variables to their mean value.

Table 5: Representative multiplier effects under different scenarios

(preferred model: Table.4, column 2)

Study characteristics (other factors held constant)	**Combined effect**	**t-stat**	**p-value**	**CI**
Study estimated by SEE (incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)	1.549	3.67	0.001	0.70;2.39
Study estimated by VAR	2.423	4.75	0.000	1.4;3.44

[10]Firstly, if EXP is excluded from the model, RECC remains insignificant while FINCRICIS turns out to be statistically significant and positive. Second, if both EXP and RECC are excluded FINCRISIS slightly gains in significance. Third, if RECC is excluded, EXP remains significant while FINCRISIS results insignificant.

(incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)				
Fiscal impulse is CONSUMPTION (incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)	3.041	6.09	0.000	2.03;4.04
Fiscal impulse is INVESTMENT (incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)	2.995	5.83	0.000	1.96;4.02
Fiscal impulse is MILITARY SPENDING (incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)	2.294	3.81	0.000	1.08;3.5
Fiscal impulse is TAX SHOCK (incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)	1.934	4.17	0.000	1;2.86
Fiscal impulse is PUBLIC EMPLOYMENT (incr=0; tempor=0; transit=0; quart=0; horiz=mean; primary data=mean)	2.508	3.98	0.000	1.24;3.76
Study estimated by SEE, fiscal impulse is NOTSPE, (incr=1; tempor=1; exp=1; quart=1; japan=1)	-1.504	-2.64	0.011	-2.64;-0.36
Study estimated by SEE, fiscal impulse is TAX, (incr=1; tempor=1; exp=1; quart=1; japan=1)	-1.957	-3.38	0.001	-3.11;-0.79
Study estimated by SEE, fiscal impulse is TAX, (incr=1; tempor=1; transit=0; quart=1; horiz=mean; primary data=mean)	-0.627	-2.41	0.02	-1.15;-0.10
Study estimated by VAR, fiscal impulse is TAX, (incr=1; tempor=1; transit=0; quart=1; horiz=mean; primary data=mean)	0.25	2.01	0.049	0.001;0.49
Study estimated by VAR, fiscal impulse is TAX, (incr=0; tempor=1; exp=1; quart=1; horiz=mean; primary data=mean)	-0.003	-0.02	0.459	-0.34;0.76

Note: the results were obtained using Stata 'lincom' command. See Appendix A.14 for details.

After controlling for publication bias, the representative effect of a study, using a SEE model for a negative permanent unspecified government expenditure shock estimated by a conventional model using annual data and controlling for structural characteristics of the country, is estimated to be about 1.55 and significantly different from zero. A study with the same characteristics but estimated by a VAR provides a representative effect of 2.43. If a study uses VAR model to estimate the permanent decrease in consumption for an advanced economy using annual data and controlling for country characteristics, it yields a multiplier of 3.04. Similar study estimated for public investment and public employment provides a multiplier of 2.99 and 2.51, respectively. If the fiscal impulse is military spending or tax cut, the representative effects are 2.29 and 1.93, respectively. Studies estimated by VAR, investigating a temporary increase in taxes in advanced economies using quarterly data yields a multiplier effect of 0.25. A similar study estimated by SEE yields a negative multiplier of -0.62. A study for Japan estimated by SEE for a temporary increase of unspecified government expenditure, using quarterly data and assuming Japan economy is in expansion yields a negative multiplier of -1.50. A similar study estimating the effect of tax increase yields a negative multiplier of -1.95. These results suggest that multipliers under different scenarios can take a wide spectrum of values ranging from negative to positive values. Therefore considering the 'overall multiplier effect'[11](0.68; p=0.00) is meaningless and appears to be completely misleading if used for policy analysis.

Conclusion

In this study, as a complement to the narrative literature review, a MRA analysis is employed on a set of 65 empirical VAR and SEE studies, in order to review the literature with statistical criteria and to provide explanations about the heterogeneity of empirical results as well as investigate possible publication selection bias in the literature.

[11] See Appendix A.18 for the calculation of the 'overall' multiplier.

The results of our analysis suggest that the type of model class is important in explaining the heterogeneity of reported multipliers. VAR models estimates differ significantly from SEE estimates. Regarding the type of fiscal shock, public consumption appears to be the most effective impulse while tax shocks the least effective impulse. In line with a priori expectations, permanent shocks yield higher multipliers and multipliers estimated for a longer horizon are higher than impact multipliers. Quarterly data yield lower multipliers compared to annual data. Fiscal policy in transition countries appears to be more effective compared to advanced economies, contrary to a priori expectations, however the results are not stable across different specifications. In the VAR analysis, different identification strategies estimate significantly different multipliers. In line with theoretical predictions, the reported multipliers are smaller if estimated for open economies and economies in expansion and bigger for relatively more closed economies and economies experiencing a financial crisis.

Structural characteristics appear to explain less of the variation of multiplier estimates. From the labour market characteristics, only replacement rates appear to be a significant determinant of the multiplier effect. The degree of openness is an important factor while variables accounting for the indebtedness of the economy, financial development and monetary policy appear to insignificant factors in determining multipliers, contrary to theoretical predictions.

Investigating publication selection bias, the funnel plot analysis as well as the formal FAT-PET test applied by using sample size as a proxy for precision, suggests that the literature on fiscal multipliers is infected by publication bias. Controlling for publication bias is thus an important feature of this MRA.

Conclusively, our findings suggest that the heterogeneity of the reported multipliers mostly arises from the study characteristics: data settings and methods of estimation, while the effect of structural characteristics is mixed. However, the degree of openness and labour market characteristics appear to be important factors that should be considered while estimating fiscal multipliers. In particular, these findings support the idea to augment the VAR in primary study with labour market institutional characteristics.