Abstract

This paper reconciles seemingly disparate estimates of multipliers within a unified and state-contingent framework. We achieve identification of causal effects with new propensity-score based methods for time series data. Using this novel approach, we show that austerity is always a drag on growth, and especially so in depressed economies: a one percent of GDP fiscal consolidation translates into a loss of 4 percent of real GDP over five years when implemented in a slump, rather than just 1 percent in a boom.

In this paper we delve into the experimental toolkit so as to re-examine a key issue for macroeconomics, the need to ensure treatments are somehow re-randomized in non-experimental data. We do this in the context of the foremost academic and policy dispute of the day—the effects of fiscal policy shocks on output (see, in particular, Alesina and Ardagna 2010; Guajardo, Leigh, and Pescatori 2014).

Following a new and arguably more promising direction, we take a third fork on the road to identification based on the Rubin Causal Model. This approach has the attractive features of being semiparametric (and hence flexible with respect to the functional form), providing better control for observables, and offering a more reliable alternative when the putative instrumental variables for policy action are themselves possibly endogenous. Tests of instrument validity are well-known to have low power (see, e.g. Cameron and Trivedi 2005) but, more importantly, formal testing is not an option when we are in the case of exact identification.

We arrive at this conclusion by carefully constructing an encompassing framework that allows us to evaluate the type of approach followed by several recent papers in the literature (to be discussed in detail shortly) to improve comparability with the methods we introduce in this paper. In addition to accommodating existing methods, the framework allows us to address the identification concerns we uncover via the application of an estimator from the family of “doubly robust” augmented inversepropensity-score weighted regression adjustment methods (Robins, Rotnizky, and Zhao 1994; Robins 1999; Scharfstein, Rotnitzky, and Robins 1999; Hirano, Imbens, and Ridder 2003; Lunceford and Davidian 2004; Imbens 2004; Glynn and Quinn 2010).

1.

Second, we use (Jorda` 2005) local projections (LPs), rather than structural vector auto regressions (SVARs). The reason is that, among other advantages that we will discuss momentarily, LPs are a convenient pedestal on which all extensions of existing estimation methods can rest. The unified framework provides the reader a way to compare the results across a set of nested estimation strategies. LPs provide a flexible semi-parametric regression control strategy to estimate dynamic multipliers and include, as a special case, impulse responses calculated with an SVAR. LPs accommodate possibly nonlinear, or state-dependent responses easily, and indeed we find that the effects of fiscal policy can be very different in the boom and the slump, as emphasized by Keynes in the 1930s. State-dependent multipliers based on LPs have been taken up in some very recent papers (Auerbach and Gorodnichenko 2012, 2013, for the US and OECD; Owyang, Ramey, and Zubairy 2013, for the US and Canada). Other recent papers on state-dependent multipliers, using various measures of slack, include Barro and Redlick (2011) and Nakamura and Steinsson (2014). For a critical survey see Parker (2011). Long ago, Perotti (1999) explored the idea of “expansionary austerity” with state-dependent multipliers. We calculate the impact of fiscal policy shocks based on LPs using the AA measure of policy, the change in the cyclically-adjusted primary balance (d.CAPB). When we restrict attention to “large” shocks (changes in CAPB larger in magnitude than 1.5% of GDP, which is the benchmark cutoff value used by AA and proposed earlier by Alesina and Perotti 1995), we replicate the “expansionary austerity” result. However, when we condition on the state of the economy, we find that this result is driven entirely by what happens during a boom. The expansionary effects of fiscal consolidation evaporate when the economy is in a slump.

Third, we then use instrumental variable (IV) estimation of the LPs to account for unobserved confounders. Specifically, we instrument the cyclically-adjusted primary balance using the IMF’s narrative measure of an exogenous fiscal consolidation in GLP. This type of “narrative-based identification” has been applied by, e.g., Mertens and Ravn (2013, 2014), Ramey and Shapiro (1998), and Romer and Romer (1989, 1997). Our IV estimation then turns out to replicate the flavor of the GLP results: austerity is contractionary, and strongly so in slumps

Fifth, in order to purge remaining allocation bias we use inverse probability weighting (IPW) estimation based on a prediction model of the narrative policy variable to estimate the LP responses. We consider the IMF narrative policy variable as a “fiscal treatment”—i.e., a binary indicator rather than a continuous variable—and we are interested in characterizing a dynamic average treatment effect (ATE). In new work, Angrist, Jorda, and Kuersteiner ( ` 2013) introduce IPW estimators in a time series context to calculate the dynamic ATE responses to policy interventions. We follow a slightly different approach using augmented regression-adjusted estimation instead, denoted AIPW, which combines inverse probability weighting with regression control and adjusts the estimator to achieve semi-parametric efficiency (see, e.g., Lunceford and Davidian 2004). Our AIPW estimator falls into the broad class of “doubly robust” estimators of which Robins, Rotnitzky, and Zhao (1994) is perhaps the earliest reference (see also Robins 1999; Scharfstein, Rotnitzky, and Robins 1999; Hirano, Imbens, and Ridder 2003; Lunceford and Davidian 2004; Imbens 2004; Glynn and Quinn 2010). The “doubly robust” property means that consistency of the estimated ATE can be proved in the special cases where either the propensity score model and/or the conditional mean is correctly specified; Monte Carlo evidence also suggests that the estimator performs better than alternatives even in more general cases to

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Fiscal multiplier, d.CAPB, OLS estimate, booms v. slumps | | | | | |  | |
|  | (1) | (2) | (3) | (4) | (5) | |
|  | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | |
| Education expend. | -0.00068 | -0.06626 | -0.09523 | -0.08122 | -0.12006 | |
| *y > 0, boom* | -0.0244 | -1.1517 | -1.5184 | -1.5972 | -2.4228 | |
| Education expend. | -0.16025 | -0.19381 | -0.18283 | -0.17263 | -0.10299 | |
| *y <= 0, slump* | -1.5571 | -1.9545 | -2.5851 | -2.4419 | -1.8977 | |
| Health expend. | 0.014831 | -0.0245 | -0.09702 | -0.10928 | -0.08907 | |
| *y > 0, boom* | 0.811 | -0.596 | -1.6686 | -2.4582 | -2.6482 | |
| Health expend. | 0.013337 | -0.04511 | -0.11222 | -0.07219 | -0.03544 | |
| *y <= 0, slump* | 0.626 | -1.1984 | -1.803 | -1.9877 | -0.716 | |
| In-kind transfers | -0.00645 | -0.03477 | -0.06014 | -0.08009 | -0.05782 | |
| *y > 0, boom* | -0.3964 | -1.0507 | -1.8925 | -2.8288 | -1.6765 | |
| In-kind transfers | -0.01876 | -0.02095 | -0.02619 | -0.02562 | -0.00345 | |
| *y <= 0, slump* | -2.7656 | -1.844 | -1.8935 | -1.8619 | -0.1802 | |
| In-cash transfers | -0.00011 | -0.04022 | -0.11223 | -0.12571 | -0.09976 | |
| *y > 0, boom* | -0.0057 | -1.1591 | -2.11 | -3.7 | -2.3876 | |
| In-cash transfers | -0.01876 | -0.02095 | -0.02619 | -0.02562 | -0.00345 | |
| *y <= 0, slump* | -2.7656 | -1.844 | -1.8935 | -1.8619 | -0.1802 | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
| Property taxes | 0.000511 | 0.004564 | 0.007615 | 0.009323 | 0.004626 | |
| *y > 0, boom* | 0.2247 | 1.6089 | 1.61 | 1.9924 | 1.3275 | |
| Property taxes | 0.002791 | 0.00816 | 0.015106 | 0.017364 | 0.009144 | |
| *y <= 0, slump* | 1.0221 | 2.7583 | 2.435 | 2.6546 | 1.6161 | |
| Indirect taxes | 0.009601 | -0.0433 | -0.10832 | -0.10088 | -0.11933 | |
| *y > 0, boom* | 0.5257 | -1.6472 | -3.1783 | -4.145 | -4.5553 | |
| Indirect taxes | 0.026934 | -0.09497 | -0.13702 | -0.11448 | -0.1331 | |
| *y <= 0, slump* | 0.6502 | -2.3676 | -2.2557 | -3.2287 | -2.943 | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
| R&D + Savings | -0.01258 | -0.02167 | -0.03951 | -0.01222 | -0.02666 | |
| *y > 0, boom* | -1.7973 | -1.4753 | -2.7197 | -0.5573 | -1.0332 | |
| R&D + Savings | -0.0148 | -0.01888 | -0.04685 | -0.04098 | -0.05858 | |
| *y <= 0, slump* | -1.3767 | -0.9552 | -2.3976 | -2.1872 | -2.8262 | |

‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fiscal multiplier, XXX, OLS estimate, booms v. slumps | | | | | |  | |  |
|  | (1) | (2) | (3) | (4) | (5) | |
|  | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | |
| Education expend. | -0.42031 | -0.45941 | -0.63161 | -0.03236 | 0.164519 | |
| *y > 0, boom* | -1.9923 | -1.5181 | -2.1726 | -0.106 | 0.5548 | |
| Education expend. | 0.04638 | 0.188081 | -0.28376 | 0.587261 | 0.56744 | |
| *y <= 0, slump* | -0.1853 | 0.6061 | -0.7348 | 1.8534 | 1.2744 | |
| Health expend. | -0.57195 | -1.06849 | -0.97031 | -0.34035 | -0.38077 | |
| *y > 0, boom* | -1.6587 | -3.2663 | -3.1653 | -1.1746 | -1.4103 | |
| Health expend. | -0.2157 | -0.50532 | -0.43922 | 0.295785 | 0.298864 | |
| *y <= 0, slump* | -0.6347 | -1.277 | -1.5918 | 0.9963 | 0.564 | |
| In-kind transfers | -0.58563 | -0.76126 | -0.47075 | -0.50763 | -0.18793 | |
| *y > 0, boom* | -2.7591 | -3.5481 | -1.5901 | -2.5431 | -0.8339 | |
| In-kind transfers | 0.146704 | -0.19295 | 0.169891 | -0.12906 | 0.396081 | |
| *y <= 0, slump* | -0.5468 | -1.1876 | 1.3499 | -0.6827 | 1.2747 | |
| In-cash transfers | -0.95255 | -1.25267 | -0.92266 | 0.830377 | -0.55081 | |
| *y > 0, boom* | -6.1854 | -5.0232 | -4.5687 | -3.851 | -1.8234 | |
| In-cash transfers | -0.71124 | -0.191 | 0.007326 | -0.03079 | -0.02675 | |
| *y <= 0, slump* | -2.2879 | -0.5983 | 0.0261 | -0.0872 | -0.0261 | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
| Property taxes | 0.012637 | 0.04311 | 0.048724 | 0.038707 | 0.033993 | |
| *y > 0, boom* | 0.5584 | 1.433 | 1.6213 | 1.1414 | 1.006 | |
| Property taxes | 0.007071 | 0.022379 | 0.044496 | 0.002223 | 0.026811 | |
| *y <= 0, slump* | 0.2354 | 0.6374 | 1.0837 | 0.0514 | 0.4623 | |
| Indirect taxes | -0.48486 | -0.54533 | -0.59608 | -0.47982 | -0.86514 | |
| *y > 0, boom* | -2.2728 | -1.6573 | -1.8594 | -2.2728 | -3.575 | |
| Indirect taxes | -0.41799 | -1.15687 | -1.18382 | -1.02618 | -1.30148 | |
| *y <= 0, slump* | -1.0942 | -4.2043 | -3.4258 | -2.8469 | -2.4154 | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
| R&D + Savings |  |  |  |  |  | |
| *y > 0, boom* |  |  |  |  |  | |
| R&D + Savings | 0.121633 | 0.080426 | -0.20576 | -0.08958 | -0.64455 | |
| *y <= 0, slump* | 0.7816 | 0.5755 | -1.2557 | -0.2553 | -2.1624 | |

Following a new and arguably more promising direction, we take a third fork on the road to identification based on the Rubin Causal Model. This approach has the attractive features of being semiparametric (and hence flexible with respect to the functional form), providing better control for observables, and offering a more reliable alternative when the putative instrumental variables for policy action are themselves possibly endogenous. Tests of instrument validity are well-known to have low power (see, e.g. Cameron and Trivedi 2005) but, more importantly, formal testing is not an option when we are in the case of exact identification.

We find that, on average, fiscal consolidations generate a drag on GDP growth. The effect is also state dependent: if a 1 percent of GDP fiscal consolidation is imposed in a slump then it results in a real GDP loss of around 4 percent over five years, rather than just 1 percent in a boom. We arrive at this conclusion by carefully constructing an encompassing framework that allows us to evaluate the type of approach followed by several recent papers in the literature (to be discussed in detail shortly) to improve comparability with the methods we introduce in this paper.

Using our estimates we compute how much of the slowdown could be attributed to the austerity program; we find it to be a very significant contribution (rising to 3.4% of GDP in 2013) and larger than official estimates. Thus, better models, with state-dependent features, could improve official fiscal policy analyses going forward.

Second, we use (Jorda` 2005) local projections (LPs), rather than structural vector auto regressions (SVARs). The reason is that, among other advantages that we will discuss momentarily, LPs are a convenient pedestal on which all extensions of existing estimation methods can rest. The unified framework provides the reader a way to compare the results across a set of nested estimation strategies. LPs provide a flexible semi-parametric regression control strategy to estimate dynamic multipliers and include, as a special case, impulse responses calculated with an SVAR. LPs accommodate possibly nonlinear, or state-dependent responses easily, and indeed we find that the effects of fiscal policy can be very different in the boom and the slump, as emphasized by Keynes in the 1930s.

Fourth, we show that the proposed IMF narrative instrumental variable has a significant forecastable element driven by plausible state variables, such as the debt-to-GDP level, the cyclical level or rate of growth of real GDP, and the lagged treatment indicator itself (since austerity programs are typically persistent, multi-year affairs).

The local projection is done from year 0, when a policy change is assumed to be announced, with the fiscal impacts first felt in year 1, consistent with the timing in GLP. The LP output forecast path is constructed out to year 5, and deviations from year 0 levels are shown, and also the sum of these deviations, or “lost output” across all of those five years

. Small consolidation packages have a small effect on output, but the estimates are imprecise.

If the IMF approach is correct and has found truly exogenous shocks to fiscal policy, then it would be a valid instrument for d.CAPB. It would also be a potentially strong instrument: the raw correlation between d.CAPB (year 1 versus year 0) and Treatment (in year 1) is 0.31, and a bivariate regression has an F-statistic of over 50; the same applies when Treatment is replaced by Total (in year 1).

The IV-based responses suggest that austerity is contractionary since the only statistically significant coefficients here have a negative sign. However, stratification by the state of the cycle shows that this result is now driven by what happens in slumps. It is only in the slump bin that we find a significant negative response of real GDP to fiscal tightening. In Table 4 we find a coefficient or “multiplier” of between -0.25 and -0.95 in years 1 to 5. Over five years the sum of these effects is -3.35∗∗, so the average loss for a 1% of GDP fiscal consolidation is to depress the output level by about -0.67% per year over this horizon.

4. Endogenous Austerity: Is the Narrative Instrument Valid? So far we have briefly replicated the current state of the literature, but this is not entirely pointless. It serves to show that the LP framework can capture different sides of the debate in a uniform empirical design, on a consistent data sample, allowing us to focus on how differences in estimation and identification assumptions lead to different results. It also shows how the LP estimation method makes it very easy to allow for nonlinearity and do a stratification of results; here we found significant variations in responses across bins designed to capture variations in the state of the economy from boom to slump.

**Statistical design**

The previous section raises concerns that the narrative IMF variable could be an invalid instrument using three different checks. The empirical strategy that we propose is based on taking triple insurance against this potential endogeneity. F

In order to facilitate the exposition we momentarily drop the cross-sectional country index in the panel. Denote, as before, yt the outcome variable of interest, the log of real GDP. In other applications yt could be a ky-dimensional vector. Let Dt denote the fiscal policy variable. Dt will now be a discrete random variable Dt ∈ {0, 1} based on the IMF narrative indicator of exogenous fiscal consolidations, although earlier it was the continuous d.CAPB variable.



That is, the treatment-control allocation is independent of potential outcomes given the variables or controls Xt . This condition does not imply that there is no effect of policy on the outcome given controls. We are simply stating that conditional on controls, policy allocation is independent of the potential outcome, whatever that might be.

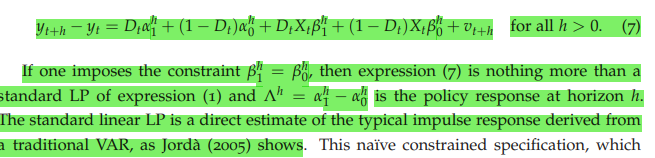
Consider the ideal randomized experiment to understand the role that the conditional independence assumption plays. The average causal effect of policy intervention on the outcome at time t + h given b.



where n1 = ∑t Dt and n0 = ∑t (1 − Dt) are the number of observations in treatment and control groups, respectively. Alternatively, the average treatment effect, Λh , could be calculated from the auxiliary regression

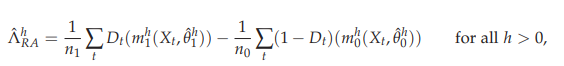


The estimator is consistent for the average treatment effect (ATE hereafter) whether or not regressors are included. Notice that the model for the outcomes is unspecified. The estimate of the ATE does not depend on specific assumptions about this model if the conditional ignobility assumption is met.



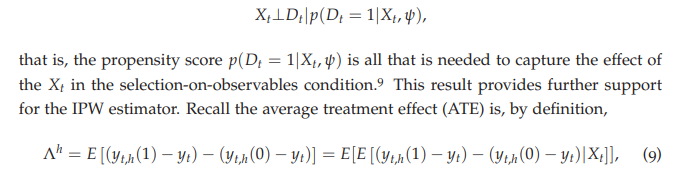
This na¨ıve constrained specification, which characterizes responses derived from a VAR, imposes two implicit assumptions. First, the effect of the controls Xt on the outcomes is assumed to be stable across the treated and control subpopulations. Second, the expected value of Xt in each subpopulation is assumed to be the same. The first assumption is potentially defensible. The economic mechanism describing the transmission of interest rates on real GDP could be the same whether or not there is a fiscal consolidation, for example. The second assumption is more difficult to defend. It is unlikely that, say, government debt levels are the same in the treated and control groups. Fiscal consolidations are often driven by high levels of debt.

These are: (1) the instrument is relevant, which appears to be the case as we discussed earlier; (2) the instrument is valid, which is untestable given just-identification and for which the analysis of the previous section raises concerns; and (3) predetermined and exogenous controls are not omitted from the specification. This latter requirement is not resolved by the use of the instrument, especially when there is substantial evidence that the controls are predictive for the instrument, as shown here.



**5.1. Re-randomization through the propensity score**

Recall that the critical assumption is the conditional ignorability or selection-on-observables condition (2). Rosenbaum and Rubin (1983) show that



Angrist and Kuersteiner (2004, 2011) refer to the predicted value from such a policy model the policy propensity score. The policy propensity score is meant to ensure the estimation of the policy response (the average treatment effect in the microeconomics parlance) is consistent under the main assumption. In addition, it acts as a dimensionreduction device. Ideally, any predictor of policy should be included, regardless of whether that predictor is a fundamental variable in a macroeconomic model. The probit results reported in Table 7 can be seen as candidate estimates of this policy propensity score. We will instead construct the policy propensity score using a richer specification that includes all the controls used in Table 6 as well.

**5.2. Regression adjustment (IPWRA) and Augmented IPW (AIPW)**

As a way to enhance robustness, researchers have derived estimators with a regression adjustment component added to the standard IPW estimator presented above. This estimator parallels that in expression (8) but using inverse probability weighting. To further enhance efficiency, the augmented IPW or AIPW estimator combines the IPW and IPWRA estimators in a manner to be discussed shortly.

The inverse propensity-score weighted estimator with regression adjustment (IPWRA)

The intuition behind the estimator is to use the regression model as a way to “predict” the unobserved potential outcomes. Consistency of the estimated ATE only requires either the conditional mean model or the propensity score model to be correctly specified.

**5.3. Intuition**

Although these techniques are relatively new to macroeconomics, matching estimators using inverse propensity score weighting have been frequently implemented in applied. microeconomics with cross-sectional data. Matching methods more generally constitute a benchmark within the medical research literature when trials are suspected of being contaminated by allocation bias. The provenance of the particular inverse propensityscore weighting method we employ is thus well established.

Weighting by the inverse of the propensity score shifts weight away from the oversampled toward the undersampled region of the distribution. This shift of probability mass reconstructs the appropriate frequency weights of the underlying true distribution of outcomes under treatment and control so that the means estimated from each subpopulation are no longer biased and their difference is an unbiased estimate of the ATE.

**As good as random**

Instead of assuming that the treatment is randomly assigned, we assume that the treatment is as good as randomly assigned after conditioning on covariates Formally, this assumption is known as conditional independence Even more formally, we only need conditional mean independence which says that after conditioning on covariates, the treatment does not affect the means of the potential outcomes