**Abstract**

We achieve identification of causal effects with new propensity-score based methods, using this approach, we show that fiscal multipliers have different effects on income inequality, 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 research also re-examine a key issue for macroeconomics and statistical experiments, the need to ensure treatments groups are somehow re-randomized in non-experimental data. We do this in the context of studying the effects of fiscal policy shocks on income inequality (see, in particular, Alesina and Ardagna 2010; Guajardo, Leigh, and Pescatori 2014).

We followed a promising direction, based on the Rubin Causal Model. Considering semiparametric features are flexible with respect to its 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.

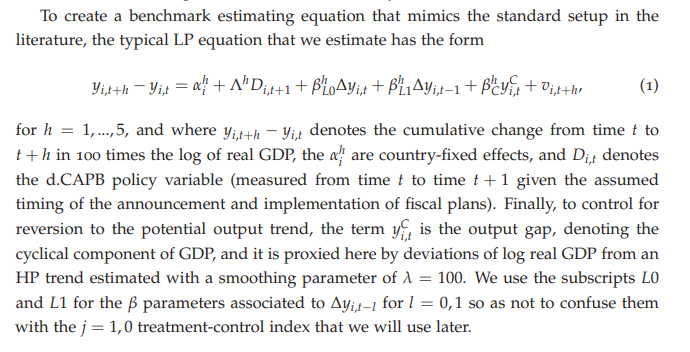
We evaluate our model following several recent papers in the literature to improve comparability with the statistical methods we introduce in this paper. In addition, this framework allows us to address the identification concerns via the application of an augmented inverse propensity-score weighted 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 hereafter), rather than structural vector auto regressions (SVARs). The reason is that, LPs are a convenient pedestal on which all extensions of existing estimation methods can rest. The unified framework provides 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 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. We calculate the impact of fiscal policy shocks based on LPs using a set of social transfers, government expenditure and taxing. We also restrict our attention to “large” shocks (changes in GDP) to introduce and observe the “booms and bust” results. 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.

Fifth, in order to purge remaining allocation bias, we use inverse probability weighting (IPW) estimation based on a prediction model of the GDP shocks variable to estimate the LP responses, we consider these policy variables 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). We follow an augmented regression-adjusted estimation instead, denoted AIPW, which combines inverse probability weighting with regression control adjusting 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). 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.

2.



Our choice of using the HP filter with λ = 100 was justified by a series of experiments undertaken with US postwar data (from FRED) which showed that a relatively high smoothing parameter was needed if the proposed proxy series (HP filtered log real GDP) was to come close to matching the official CBO output gap series. We also replicated this type of analysis using a bandpass filter tuned to various frequencies, and the conclusions were very similar. That is to say, we found that the conventional filter frequencies typically used in the business cycle literature are too low to provide a good match with the output gap, which is what we want in our model so as to control for reversion to trend. These experiments are available from the authors upon request.

Note that this partition is meant to provide a more granular statistical summary of the main features of the data. We are not arguing whether or not a boom or a slump is more likely under a particular choice of fiscal policy or another.

The results are reasonable and consistent with the literature, and particularly the GLP replication of the AA-type results. The OLS estimates suggest that fiscal austerity is expansionary, since the only statistically significant coefficients are ones that have a positive sign. However, our stratification of the results by the state of the cycle at time 0 brings out a new insight, and shows that this result is entirely driven by what happens in booms. It is only in the boom that we find a significant positive response of real GDP to fiscal tightening, with a coefficient or “multiplier” (the more general usage of the term, which we follow in the remainder of the paper) of nearly 0.25 in years 1 and 2. Over 5 years the sum of these effects is small, near 0.15. In the slump, the estimate of the policy response is not statistically different from zero and in many cases it is negative.

Table 3 –insert gere—

To bring this IMF approach into our framework, and consistent with our OLS replication of the AA results above, we present in Tables 3, and 4 our IV estimates which make use of the IMF narrative variable. We reestimate expression (1) using the IMF dates of fiscal consolidations as both binary and continuous instruments. This approach is parallel to the approach in Mertens and Ravn (2013, 2014) for the US and based on Stock and Watson (2012).

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).

We begin by reestimating the full sample specification reported in the top panel of Table 1 using instrumental variables in two ways. First, we use the IMF narrative variables on dates of fiscal consolidation as a binary instrument (first row). Second, for a continuous IV we use the size of the consolidation identified by the IMF (second row). The results are reported in Table 3. Strikingly, the message here completely overturns the findings in Table 1. This is of course a well-known problem, consistent with the pronounced divergence between the AA and GLP results. Fiscal consolidation is unambiguously contractionary. Using the sum of coefficients reported in column (6) of Table 3, for every 1% in fiscal consolidation, the path of real GDP is pushed down by over 0.57 percent each year on average over the five subsequent years. This result is not sensitive to whether we use the binary or continuous instrument.

Before drawing any conclusions, we evaluate whether the IMF narrative variable might be a legitimate instrument. Have we identified the causal effect of fiscal consolidations on output? We cannot formally test the validity of the IMF narrative instrument since the LPs are just identified. However, if the IMF’s narrative variable can be predicted by excluded controls, and those controls are correlated with the outcome, at a minimum the excluded controls should be added to the regression. At worst, predictability points to having failed to resolve the allocation bias in our estimates—episodes of consolidation identified by the IMF might be simply an endogenous response by the fiscal authority. This possible shortcoming of the “narrative identification” strategy has been noted before in the context of monetary policy (Leeper 1997) and we have the same concern here.

Si me animo Tabla 5 (ks test in R), solo se menciona 1 vez.

Tabla 6: <https://bookdown.org/ccolonescu/RPoE4/RPoE.pdf> (pg. 75)

Tabla 7, Hacer

Next we check for another condition: Do excluded controls predict fiscal consolidations? Table 7 asks whether variation in the IMF binary treatment variable identified by GLP can be predicted. The results indicate that we have a reasonable basis for this concern.

Table 7 shows in column (1) that treatment is more likely, as expected, when public debt to GDP is high: the coefficient is positive, meaning that governments tend to pursue austerity when debt has run up. In column (2) we add y C (the output gap) and the growth rate of y to further condition on the state of the economy: when the economy is growing below potential, there is an increase in the likelihood of consolidation.

Thus, the act of engaging in pro-cyclical fiscal policy is not a new-fangled craze but more of a chronic tendency in advanced countries. Finally, columns (3) and (4) add the lag of the dependent variable Treatment and this has a highly significant coefficient: as we know from the raw data series generated by the IMF study, the fiscal consolidation episodes are typically long, drawn-out affairs, so once such a program is started it tends to run for several years. Being in treatment today is thus a good predictor of being in treatment tomorrow. In these last two columns the lagged growth rate rather than the cyclical level of output emerges as the slightly better predictor of treatment.

Further confirmation of the predictive ability of these treatment regressions is provided by the AUC statistic.7 The AUC is commonly used in biostatistics and machine learning to evaluate classification ability (see, e.g. Jorda and Taylor ` 2011). Under the null that the covariates have no classification ability, AUC = 0.5. Perfect classification ability corresponds to AUC = 1. The AUC has an approximate Gaussian distribution in large samples. Table 7 measures the classification ability of each specification. The AUC statistics show that the probits have very good predictive ability, with AUC around 0.65 when lagged treatment is omitted (Column 2), and over 0.8 when lagged treatment is included (Columns 3 and 4). The AUCs are all significantly different from 0.5.

The key lesson from Table 7 is simply that the IMF variable has a significant forecastable component.8 The question, then, is how to deal with the problem of potentially endogenous instruments. The remainder of this paper provides one answer.

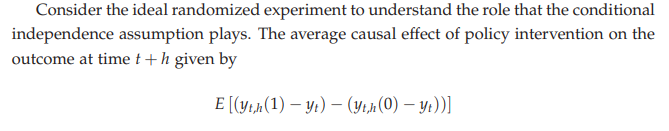
**Statistical design**

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. First, we take the episodes of consolidation from the IMF narrative variable as the subset of all consolidation episodes that are a candidate for random allocation. Think of it as a pseudo-IV step. Second, we include the extended set of covariates from Tables 6 and 7 and add them as right-hand side variables in the LP of expression 1. Third, we use inverse propensity score weighting on this LP to re-randomize allocation of the IMF fiscal consolidation events.

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.

The methods that we present next can be extended to settings in which the policy variable takes on a small number of discrete values. Next we allow for a kw-dimensional vector of variables, wt that are not included in the vector yt , but which could be relevant predictors of the policy variable Dt . Finally, denote Xt the rich conditioning set given by ∆yt−1, ∆yt−2, ...; Dt−1, .Dt−2, ...; and wt.

The causal effect of a policy intervention is defined as the unobservable random variable given by the difference (yt,h (1) − yt) − (yt,h (0) − yt). Notice that yt is only used to benchmark the cumulative change and it is observed at time t. We assume that the parameters of the policy function do not change.



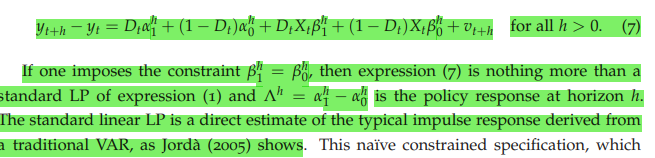


where n1 = ∑t Dt and n0 = ∑t (1 − Dt) are the number of observations in treatment and control groups, respectively. Alternatively, the average treatment effect, Λh , could be calculated from the auxiliary regression



Even when data are randomly allocated across the treatment and control subpopulations, it would be natural to condition on the Xt to adjust for small-sample differences in subpopulation characteristics and therefore to gain in efficiency. 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.

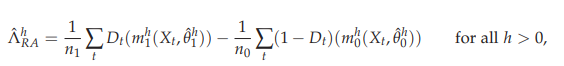
Assume that a linear regression control strategy suffices to do the appropriate conditioning for the Xt and hence obtain a consistent estimate of E[yt+h − yt |Dt , Xt ]. This is a big assumption that we relax later on in the paper. Note this is the assumption of studies based on VARs where identification does not rely on external information. Then the average causal effect of a policy intervention on the outcome variable at time t + h in the maintained example, can be calculated by expanding expression (4) with



This naïve 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.

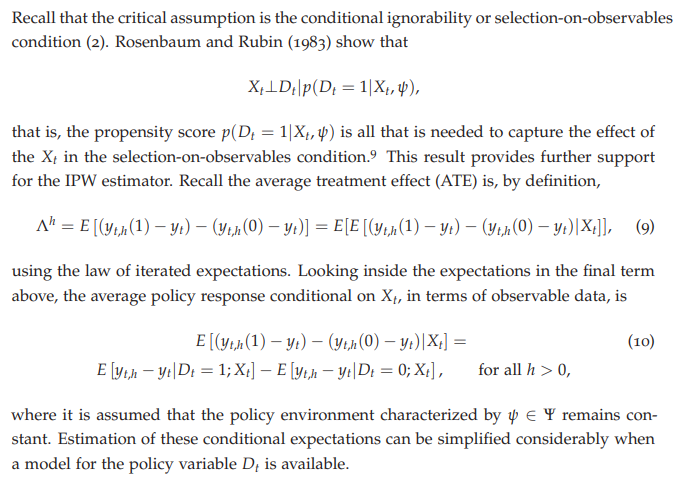
When instruments are available one can further achieve identification using instrumental variable methods as in Stock and Watson (2012) and Mertens and Ravn (2013, 2014). We have shown above how IV methods can be used with the LP approach in a more natural way. However, it is important to recall several features required to resolve the identification puzzle. 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.

More generally, if we do not impose the implicit assumptions of the na¨ıve LP specification, the analogous representation to the group means expression (3) is

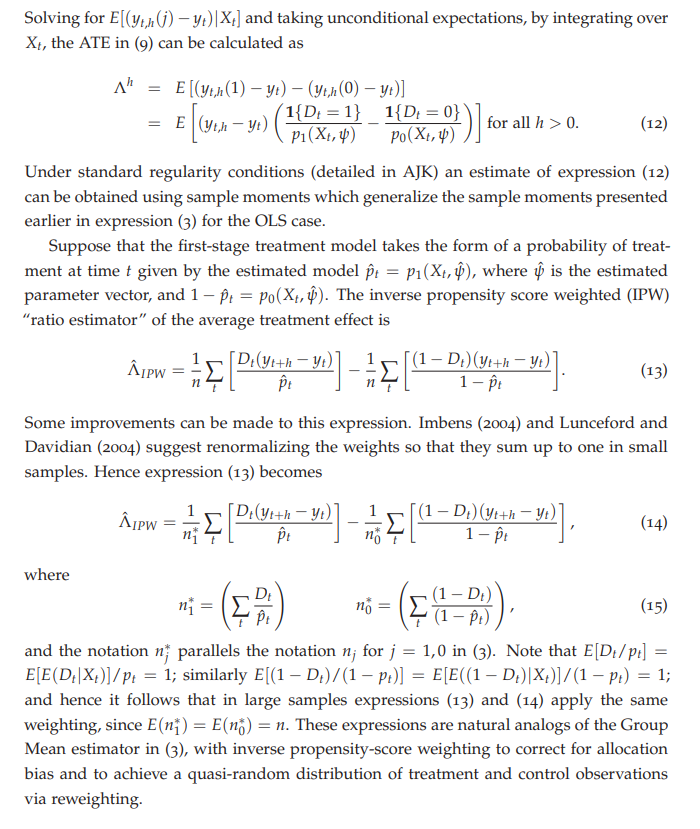


where mh j (.) is a generic specification of the conditional mean of (yt+h − yt) in each subpopulation j = 1, 0 and θ h j = (α h j β h j ) ′ for the regression example in (7). The n1 and n0 have been defined earlier. Note that this more general form of regression adjustment allows the conditional means to be different for the treated and control subpopulations and allows their effect on the outcome to differ as well.

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



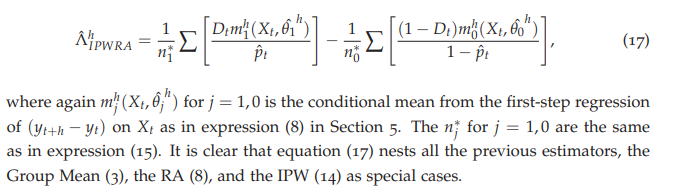
Angrist and Kuersteiner (2004, 2011) refer to the predicted value from such a policy model the policy propensity score. The policy propensity score is meant to ensure the estimation of the policy response (the average treatment effect in the microeconomics parlance) is consistent under the main assumption. In addition, it acts as a dimension reduction 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.

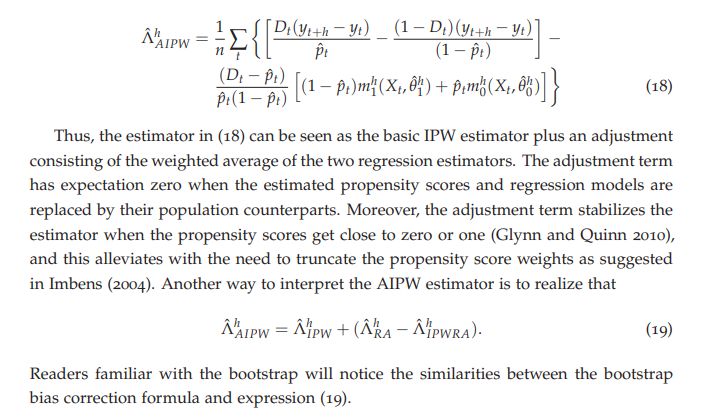


**5.2. Regression adjustment (IPWRA) and Augmented IPW (AIPW)**

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)





**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.

5.4. What We Do

The next section reports the results of applying the AIPW estimator (18) to measure the average treatment effect of fiscal consolidations as a counterpoint to the conventional OLS and IV results reported earlier. As a way to understand where the differences come from, we first implement the AIPW estimator by restricting the parameters of the regression (based on LPs) to be the same in the treated and control subpopulations, as is implicit in the OLS and IV approaches. Under that constraint, the results from the AIPW estimator are close to the IV results seen earlier. Next we allow for the parameters to vary across subpopulations, adhering to the way expression (18) is typically applied in the policy evaluation literature. These results deliver the same qualitative implication of contractionary austerity, but show that the effects of consolidations are quantitatively even more painful.

6. Contractionary Austerity Revisited: Estimates of the Average Effect of Fiscal Consolidations

This section presents AIPW estimates of the ATE of fiscal consolidations. Following standard procedures, the propensity score used here is based on a saturated probit model that extends the set of controls used in Table 7 with the current and lagged values of the controls in Table 6.

Figure 2 provides smooth kernel density estimates of the distribution of the propensity score for the treated and control units to check for overlap. One way to think of overlap is to consider what overlap would be in the ideal RCT. The empirical distributions of the propensity score for treated and control units would be uniform and identical to each other. At the other extreme, suppose that treatment is allocated mechanically on the basis of controls. Then the distribution of treated units would spike at one and be zero elsewhere, and the distribution of control units would spike at zero and be zero elsewhere.

However, the AIPW has the property that high weights in the IPW are compensated at the same rate by the augmentation term. Experiments not reported here indicate that this is indeed what happens in practice and that truncation is unnecessary in our application (see Appendix A.3).

Table 8 is organized into two rows. The first row reports the results based on imposing the restriction θ h 1 = θ h 0 , the usual implicit restriction used without hesitation in the macro-VAR empirical literature and the same restriction we imposed in reporting the results of Tables 1 and 3. The second row reports the results that do not impose the θ h 1 = θ h 0 restriction. The results are qualitatively similar to those reported in Table 3 in that we still find that austerity is contractionary. However, the estimated impacts of fiscal consolidations on output are now even bigger.

Recall that according to the IV estimates, the accumulated loss over five years was -2.94∗∗∗. This would imply an average annual real GDP loss of about 0.59% of GDP per 1% of fiscal consolidation over each of the 5 years. Here our AIPW estimate with unrestricted coefficients has a sum effect of -3.61∗∗∗ over 5 years. This would imply an average annual real GDP loss of about 0.74% of GDP per 1% of fiscal consolidation over each of the 5 years (using a 1/0.97 rescaling factor). Thus the implied output losses due to austerity are about 20% larger under our AIPW estimation than with IV estimation.

Table 9 presents these AIPW estimates based on the same saturated policy propensity score probit model described earlier. These results show that in a boom a fiscal consolidation has on average a small, negative, but imprecisely estimated effect. The first row of the table indicates that the accumulated loss over five years is -1.80 percent of GDP. In a slump, the results are about three times as strong and highly statistically significant: over five years, the accumulated loss is -3.54∗∗ percent of GDP, as shown in the second row of the table. Scaling these effects for the average treatment size (0.97 percent of GDP) the average loss per 1% fiscal consolidation is 0.37% of GDP per year over the five-year window in booms, and 0.73% of GDP per year in slumps.

Summing up our LP results, we always find more adverse paths when austerity is imposed in slumps rather than in booms, but there are sometimes big differences across specifications.

. Given the data, there would be too few observations to obtain robust results (and in some cases, insufficient data to estimate the desired effects). Fortunately, as we have discussed earlier, fiscal consolidations typically average about one percent relative to GDP with a tight range of variation, which greatly facilitates the interpretability of our findings

Figure 3 displays the coefficients reported in Table 9, with appropriate rescaling in the case of AIPW to allow for the average treatment size, to show the dynamic ATE impacts of fiscal consolidations in graphical form and compares them with the responses obtained using the IV coefficient estimates which were reported earlier in Table 4. Our results underscore that austerity tends to be painful, but that timing matters: the least painful fiscal consolidations, from a growth and hence budgetary perspective, will tend to be those launched from a position of strength, that is, in the boom not the slump. This would seem to require moderately wise policymaking and/or fiscal regimes (councils, rules, etc.), not to mention an ability to stay below any debt limit so as to maintain capital market access to permit smoothing.

7. COUNTERFACTUAL

8. CONCLUSIONS

Rather, the main contribution is to harmonize dissonant views into a unified framework where the merits of each approach can be properly evaluated. The effect of fiscal consolidation on macroeconomic outcomes is ultimately an empirical question. In the absence of randomized controlled trials, we have to rely on observational data. And to measure the causal effect of fiscal consolidations on growth, it is critical that identification assumptions be properly evaluated and that empirical methods be suitably adjusted to the demands of the data.

This result provides some measure of comfort on the potential validity of the instrument. Our analysis suggests even larger austerity impacts than the IMF study when the economy is growing below its long-run trend, however. This is likely a result of correcting attenuation bias due to the omitted predictors of fiscal consolidation and the re-randomization methods that we use. Generally, in the slump, austerity prolongs the pain, much more so than in the boom. It appears that Keynes was right after all.