DISCUSSION PAPER SERIES

DP15330
(v. 2)

SHOULD WE TRUST CROSS SECTIONAL MULTIPLIER ESTIMATES?

Fabio Canova

INTERNATIONAL MACROECONOMICS AND FINANCE

MONETARY ECONOMICS AND FLUCTUATIONS

PUBLIC ECONOMICS

ISSN 0265-8003

SHOULD WE TRUST CROSS SECTIONAL MULTIPLIER ESTIMATES?

Fabio Canova

Discussion Paper DP15330
First Published 01 October 2020
This Revision 03 October 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SHOULD WE TRUST CROSS SECTIONAL MULTIPLIER ESTIMATES?

Abstract

I examine the properties of cross sectional estimates of multipliers, elasticities, or pass-throughs when the data is generated by a conventional multi-unit time series specification. A number of important biases plague estimates; the most relevant one occurs when the cross section is not dynamic homogenous. I suggest methods that can deal with this problem and show the magnitude of the biases cross sectional estimators display in an experimental setting. I contrast average time series and average cross sectional estimates of local fiscal multipliers for US states.

JEL Classification: E0, H6, H7

Keywords: Cross sectional methods, dynamic heterogeneity, partial pooling, fiscal multipliers, Monetary pass-through

Fabio Canova - fabio.canova@eui.eu
Norwegian Business School and CEPR

Acknowledgements
Thanks to Christian Matthes and Evi Pappa for useful comments. Part of this work was conducted while the author held a Santander Chair of Excellence at UC3M, Madrid.

Should we trust cross-sectional multiplier estimates?

Fabio Canova *
Norwegian Business School, CAMP and CEPR

October 3, 2020

Abstract

I examine the properties of cross sectional estimates of multipliers, elasticities, or pass-throughs when the data is generated by a conventional spatial time series specification. A number of important biases plague estimates; the most relevant one occurs when the cross-section is dynamic heterogeneous. I suggest methods that work well in this situation and show the magnitude of the biases cross-sectional estimators display in an experimental setting. I contrast average time series and average cross-sectional estimates of local fiscal multipliers for US states.

Key words: Cross sectional methods, dynamic heterogeneity, partial pooling, fiscal multipliers, monetary pass-through.

JEL Classification: E3, E6, C3, H7.

*Thanks to Christian Matthes and Evi Pappa for useful comments. Part of this work was conducted while the author held a Santander Chair of Excellence at UC3M, Madrid.

1 Introduction

It is common in the applied literature to examine how output or employment dynamically respond to a government expenditure increase, see e.g., Nakamura and Steinnson [2014],Chodorow-Reich [2019], or to analyze how banks' lending policy dynamically evolves in response to a monetary policy change, see e.g., Jimenez, Ongena, Peydro', and Saurina [2014], Chakraborty, Goldstein, and MacKinlay [2020], using "cross-sectional" methods.

Researchers employing these methods often highlight the advantages of the approach relative to a conventional time series methodology. For example, it is often emphasized that the latter requires exogenous conditioning variables but, typically, both fiscal and monetary variables adjust to the trajectory of the economy. Thus, one needs to eliminate their endogenous variations prior to the computation of dynamic effects. To purge fiscal expenditure variables from endogenous variations one could focus attention, for instance, on war financing or employing narrative information, see Ramey [2019] for a review; to eliminate endogenous variations in monetary policy variables, one can employ, instead, financial market movements or other high frequency proxies, see Gertler and Karadi [2015].

This literature also stresses that policy variables tend to move together. For example, if government expenditure increases, taxes, debt, or the term structure of nominal rates may be simultaneously affected. Thus, the quantities one computes with a time series methodology have to be interpreted as average dynamic responses *across* different policy regimes. A cross-sectional approach is said avoid both problems because the use of a time effect in the regression eliminates endogeneity and absorbs, to a large extent, the comovements present in other policy variables. In addition, meaningful estimates can be obtained by exploiting random variations across units, that are easier to justify from an economic point of view than random changes across time.

The contribution. This paper highlights the problems that mechanical application of cross sectional methods display when computing macroeconomic objects in spatial settings. When the dynamic evolution of the cross-section is homogeneous, that is, when unit dynamically comove in response to the policy change, the approach works but under a set of stringent assumptions; when these are violated, distortions in the magnitude and the significance of propagation effects emerge. When the evolution of the cross-section is, instead, dynamically heterogeneous, as it is generally the case with cross country or cross regional macroeconomic data, the approach fails to consistently measure the average dynamic effects of interest.

I discuss two alternative procedures, one which is appropriate when time time series dimension T is large and one which is appropriate when T is small, both of which may provide superior estimates of the average dynamic effects of a policy change when dynamic heterogeneity is important.

I run a Monte Carlo experiment to measure the magnitude of the distortions different approaches display in different situations. I confirm that the alternative methods have a hedge whenever the dynamic evolution of the cross-section is heterogeneous. I also demonstrate they work well and are competitive with cross sectional approaches with a dynamically homogeneous cross-section.

I compute the average local fiscal multiplier for US states using different methodologies. When cross-sectional methods are employed the average multiplier is estimated to be zero, if not negative; with one of the alternative approaches, it is statistically indistinguishable from one, making policy conclusions quite different. I show that differences emerge because the dynamic evolution of gross state product is far from homogeneous.

The rest of the paper is organized as follows. Section 2 briefly describes cross-sectional approaches. Section 3 discusses the assumptions needed for the methodology to work when a conventional spatial data generating process is assumed. Section 4 shows the problems faced by the approach when dynamic heterogeneity is present. Section 5 provides two alternative procedures. Section 6 computes average local fiscal multipliers for US states. Section 7 concludes.

2 The cross-sectional methodology in a nut-shell

The methodology advocated to measure the average dynamic effects of a fiscal or a monetary policy change is simple. Assume that a vector of time series for unit $i = 1, \ldots, N$ is available and that the sample size T is common to all units; if not let $T = min_i T_i$. Rather than exploiting time series variations to estimate the dynamic causal effects of an unexpected policy change, unit by unit, as it is done, for instance, in the comparative VAR literature, see Kim and Roubini [2000], the methodology estimates average dynamic effects exploiting cross-unit variations at a particular point in time. Thus, the typical regression one runs is:

$$Y_{i,t+h} - Y_{i,t} = \alpha_{h,t} + \beta_h P_{i,t} + \gamma_h X_{i,t,h} + e_{i,t+h} \tag{1}$$

where $P_{i,t}$ is the policy variable and $X_{i,t,h}$ is a vector of controls. The dependent variable is the $h = 1, \ldots, H$ period change in the endogenous variable $Y_{i,t}$ for the unit i, a scalar for each t and i, $\alpha_{h,t}$, is a time effect, h is the horizon of interest, and β_h is the multiplier (elasticity, pass-through) at horizon h. It is a cross-sectional object because, given $\alpha_{h,t}$, its identification comes from variations in the policy variable across i, for a fixed t. Note that in equation (1), there is a β_h for each horizon h and that the specification permits control variables $X_{i,t,h}$ to change with h, even though this option is seldomly used in the cross-sectional literature.

Variations on equation (1) are possible. For example, Nakamura and Steinnson [2014] use $P_{i,t+h} - P_{i,t}$ as the policy variable; per-capita variables may be employed; and fixed effects may be included if $Y_{i,t}$ is not demeaned. In addition, when computing fiscal multipliers, $Y_{i,t+h} - Y_{i,t}$ and $P_{i,t}$ may be scaled by $Y_{i,t}$, see Dupour and Guerrero [2017] or by its trend component, as suggested in Ramey and Zubairy [2018]. The interpretation of β_h differs, but the estimation approach is the same for all specifications.

When (1) is the data generating process, consistent estimation of β_h requires, conditional on $X_{i,t}$, that variations in $P_{i,t}$ at each t are uncorrelated with the trajectory of the $Y_{i,t+h} - Y_{i,t}$ vector across units. Because $P_{i,t}$ is not necessarily strictly exogenous, even after a time effect is included, OLS estimation of (1) is invalid. Hence, one typically looks for one or more instruments $Z_{i,t,h}$ satisfying the standard orthogonality ($E_t[Z_{i,t,h} e_{i,t+h} | X_{i,t}] = 0$) and relevance $[Z_{i,t,h} P_{i,t} | X_{i,t}] > 0$) conditions and employs a IV (2SLS) methodology. Note that the instruments $Z_{i,t,h}$ may also be horizon dependent.

Often researchers aggregate (1) across horizons to estimate cummulative dynamic effects using the regression:

$$\sum_{h=1}^{H} (Y_{i,t+h} - Y_{i,t}) = \alpha_t + \beta P_{i,t} + \gamma X_{i,t} + \sum_{h=1}^{H} e_{i,t+h} \tag{2}$$

where $\alpha_t = \sum_{h=1}^{H} \alpha_{h,t}, \beta = \sum_{h=1}^{H} \beta_h, \gamma = \sum_{h=1}^{H} \gamma_h$. While β_h gives the dynamic response of $Y_{i,t+h} - Y_{i,t}$ to a change in $P_{i,t}$, β represents the cumulative change in the endogenous variable for horizons up to H.

3 The conditions needed for the methodology to work

Equation (1) is a legitimate empirical model, but unlikely to represent the data generating process (DGP) of spatial macroeconomic data. It neglects the presence of interdependences across the endogenous variables and units; it disregards general equilibrium effects within a unit; and restricts the dynamic response of $Y_{i,t+h} - Y_{i,t}$ to a change in $P_{i,t}$ to be the same across units.

An alternative DGP. Suppose instead that the data has been generated by a conventional time series specification:

$$y_{i,t} = a_i + b_i y_{i,t-1} + c_i p_{i,t} + d_i x_{i,t} + u_{i,t} \tag{3}$$

where a_i captures deterministic components in unit i, $y_{i,t-1}$ absorbs endogenous lag dynamics, and $p_{i,t}$ is the policy variable. Here $x_{i,t} = [y_{j,t-1}, j \neq i; w_t, w_{t-1}, \ldots]$ is a vector of controls, which may include lagged values of variables of other units, $j \neq i$, as well as global (aggregate) variables, and accounts for static and dynamic interdependences that may cause $y_{i,t}$ (and $p_{i,t}$) to comove across i. When i is small, $x_{i,t}$ can be taken to be strictly exogenous with respect to $y_{i,t}$. We also assume that, for each i, $y_{i,t}$ is a $G \times 1$ vector and that $p_{i,t}$ is a one-dimensional process; thus, if $p_{i,t}$ represents government expenditure, $y_{i,t}$ may include output, employment, debt, and deficit variables, and w_t variables controlled by monetary policy, e.g., the real rate of interest, or aggregate (global) variables.

It is worth stressing that even though $x_{i,t}$ contains lagged cross unit effects, lagged unit specific effects $y_{i,t-1}$ could be potentially important to explain the dynamics of $y_{i,t}$. For instance, if $y_{i,t}$ includes employment, its lags could matter, conditional on $p_{i,t}$, the lagged employment of other units and any aggregate variables. The error term $u_{i,t}$ is assumed to be uncorrelated over time and across units; thus it is an innovation vector and its covariance matrix is denoted by Σ_i. [1]

[1] (3) assumes that the coefficients are constant over time and the variance of the innovation process is also time invariant. Extensions to situations where the parameters depend on time (or state) are possible, but omitted here because they simply add to the problems I discuss. Also, the policy variable, $p_{i,t}$, may be allowed to be endogenously reacting to some of the components of $y_{i,t}$. Thus, to complete the specification, one would need to add another set of equations relating $p_{i,t}$ to $p_{i,t-\tau}, \tau > 1$, and to current and lagged values of $y_{i,t}$ and w_t. Because this set of equations is never explicitly described in the cross-sectional literature, I omit it from the specification.

The problems Moving (3) forward h periods and solving backward one obtains:

$$\begin{aligned} y_{i,t+h} &= a_i \sum_{m=0}^{h-1} b_i^m + b_i^h y_{i,t} + c_i(1 + b_i L + b_i^2 L^2 + \ldots b_i^{h-1} L^{h-1}) p_{i,t+h} \\ &+ d_i(1 + b_i L + b_i^2 L^2 + \ldots b_i^{h-1} L^{h-1}) x_{i,t+h} + (1 + b_i L + b_i^2 L^2 + \ldots b_i^{h-1} L^{h-1}) u_{i,t+h} \end{aligned} \quad (4)$$

Matching (4) with (1) one finds the conditions needed to insure that a cross-sectional methodology appropriately measures the average dynamic effect of a policy change at horizon h.

In particular, when estimating (1), a researcher implicitly assumes that domestic interactions do not matter when evaluating the magnitude of the policy effect (and thus $G = 1$); that the time effect $\alpha_{h,t}$ captures well the evolution of other units' endogenous variables and of the global variables; that process generating the data for each unit is non-stationary, but displays no drift ($b_i = 1, a_i = 0$); and that the impact effect of the policy and of the control variables is identical across i, ($c_i = c, d_i = d$).

Perhaps more importantly, the cross-sectional estimate of β_h captures the average effect of (discounted) cumulative changes in $p_{i,t}$ from t to $t + h$, because $P_{i,t} = (1 + b_i L + b_i^2 L^2 + \ldots + b_i^{h-1} L^{h-1}) p_{i,t+h}$. Thus, unless $p_{i,t} = \bar{p}_i$, for h=0, and $p_{i,t+h} = 0$, for $h > 0$, that is, a policy impulse is considered, the estimated effect overestimates the true effect of a policy change at time t, for each horizon h, whenever $p_{i,t}$ is positively serially correlated. Note that using $P_{i,t+h} - P_{i,t}$ is unlikely to improve the situation, unless $p_{i,t}$ is an exogenous, unit root process.

Furthermore, because the regression error in (4) has a moving average structure of order $h-1$, $e_{t+h} = (1+b_i L+b_i^2 L^2+\ldots b_i^{h-1} L^{h-1}) u_{i,t+h}$, a HAC correction is needed when computing the standard errors of the β_h estimates. The presence of a moving average structure in e_{t+h} may also jeopardize IV estimation. For example, any $Z_{i,t,h}$ which is uncorrelated with $u_{i,t+h}$ will not be necessarily uncorrelated with $e_{i,t+h}$. For the property to hold one needs to select instruments sufficiently lagged in the past, e.g. $Z_{i,t-\tau,h}, \tau > h$. While this problem could be, in part, reduced by including sufficient lags of the dependent variable in (1), this option is rarely used in the cross-sectional literature.

Finally notice that, unless lagged variables of units different from i (for example, lagged deficit of unit j when the dependent variable is output of unit i) appear in $X_{i,t}$, instruments with the right timing protocol may become invalid, as they will be correlated with the error term of (1) due to variable omission.

Because the argument I make is conditional on (3), can one it be taken to be a reasonable representation for the DGP? It is well known that any vector of time series of dimension $GN \times 1$ (the number of variables times the number of units), can be written in vector autoregressive format under standard linearity, stationarity and invertibility, see Canova [2007]. Once such a representation is obtained, (3) is a reparameterization of a the sub-block of equations belonging to unit i, which exclude the policy equation. (3) is also a reparameterization of the linear solution of a large class of equilibrium models. Thus, the DGP I consider is generic.

4 Another, more important problem

Even when $P_{i,t}$ represents an innovation in a policy variable at time t for each i, instruments are carefully selected, the controls span the space of cross-sectional interdependences, and standard errors are correctly computed, there is another issue that makes cross sectional estimates of the dynamic effects potentially misleading.

As (3) indicates, the model linking the policy variable $p_{i,t}$ to the vector of endogenous variables $y_{i,t}$ may be different for each i. In other words, a policy change of similar sign and size may have different dynamic repercussions in different units. However, to employ a cross-sectional methodology, dynamic homogeneity needs to be assumed. Alternatively, for the methodology to be meaningful, it is necessary to restrict attention to the subset of the units which are similar in their dynamic responses to policy impulses. If M dynamically homogeneous groups can be constructed, one can measure the average effect within each group, but not the average dynamic effect across groups. Note that grouping based on economic or geographical characteristics (rich vs. poor, northern vs. southern, etc.) may not be enough to generate a dynamically homogeneous groups of units. For example, Altavilla, Canova, and Ciccarelli [2020] show that the dynamic pass-through of monetary policy changes on banks' lending rates differ, even for units located in the same country, facing similar legislation, and lending to the same type of firms. Because dynamic heterogeneity is a feature of spatial macroeconomic data, neglecting it causes biases and interpretation problems.

As mentioned, it is common to include fixed effects in (1) to control for time invariant differences across units (the a_i in equation (3)). Adding fixed effects is not enough to account for the dynamic differences one should worry about: conditional on time invariant characteristics, a policy change may propagate differently in different units because, e.g., $b_i \neq b_j$. Including fixed effects would also be insufficient to account for different dynamic propagation when the policy disturbances have different cross-sectional volatilities.

Cross-sectional IV estimation with dynamic heterogeneity What happens when the dynamic evolution of the cross-section in response to policy changes is heterogeneous? It is not hard to guess that cross-sectional IV estimation fails. To provide a simple analytical illustration, I restrict attention to the situation when $b_i \neq 1$ and $b_i \neq b, \forall i$. Thus, suppose that $c_i = c, d_i = d, \forall i$ but that $b_i = b + v_i$, where b is the common component and v_i the idiosyncratic unit specific component $iid \sim (0, \sigma^2)$. This is a simple reparameterization of the DGP and has no economic or statistical implication. For example, if $N = 2$, and $b_1 = 0.98, b_2 = 0.90$, one can always choose $b = 0.94$ and set $v_1 = -0.04, v_2 = 0.04$. σ^2 plays an important role because it captures the degree of cross-sectional heterogeneity; when $\sigma^2 = 0$ the dynamic are identical across units; when $\sigma^2 = \infty$, they are completely heterogeneous. With this reparameterization, equation (4) becomes

$$\begin{aligned} y_{i,t+h} - b^h y_{i,t} &= a_i \sum_{m=0}^{h-1} b_i^m + c(1 + bL + b^2L^2 + \dots b^{h-1}L^{h-1})p_{i,t+h} \\ &+ d(1 + bL + b^2L^2 + \dots b^{h-1}L^{h-1})x_{i,t+h} + \zeta_{i,t+h} \end{aligned} \tag{5}$$

where

$$\begin{aligned} \zeta_{i,t+h} &= v_i^h y_{i,t} + c(1 + v_iL + v_i^2L^2 + \dots v_i^{h-1}L^{h-1})p_{i,t+h} + d_i(1 + v_iL + v_i^2L^2 + \dots v_i^{h-1}L^{h-1})x_{i,t+h} \\ &+ (1 + b_iL + b_i^2L^2 + \dots b_i^{h-1}L^{h-1})u_{i,t+h} \end{aligned} \tag{6}$$

To further simplify, assume that $p_{i,t} = 1, p_{i,t+h} = 0, h > 0$. Would OLS applied to (5) consistently estimate cb^h? Clearly, OLS will be invalid as the regressors of (5) are correlated with the composite error $\zeta_{i,t+h}$. Would a IV approach work? Inspection of (5) indicates that it would not. Proper instruments will be hard to find because both $x_{i,t+h}$ and $p_{i,t+h}$ appear as regressors and in the composite error.

One may argue that some form of long-run balanced response to policy changes is a basic feature of spatial economic models and that it should be imposed in estimation, even though an applied researcher is not able to prove that it holds. It turns out that cross-sectional estimation of equation (1) via OLS or IV will be hard to justify even if we assume $b_i = b, \forall i$, but allow heterogeneity in the instantaneous responses, i.e., $c_i \neq c_j$, for $i \neq j$. In this situation, the path of $y_{i,t+h}$ in response to the policy change is proportional across i but the impact effect is unit specific and the problems remain, because it is hard to find instruments correlated with the policy variable and uncorrelated with the composite error term.

Taking stock It is useful at this point to summarize the conclusions of these two sections. A cross-sectional methodology appropriately measures the average dynamic effects of a policy change if the units are dynamically homogeneous or if dynamically homogeneous groups can be created prior to estimation, but only under a set of stringent conditions. In particular, the policy variable should represent a policy impulse, otherwise the dynamic effects are incorrectly measured; the instruments should be sufficiently lagged relative to the horizon of interest, and appropriately chosen to avoid contamination from omitted variables, otherwise IV estimation fails; and standard errors should be corrected for moving average components in the regression error, otherwise the significance of the dynamic effects is overstated. When dynamic heterogeneity is instead present, cross-sectional IV estimates carry no information about the true average dynamic effects of the policy change.

5 How to proceed when dynamic heterogeneity is present?

One obvious solution to the biases I discussed in the previous section is to use machine learning algorithms to group units with similar dynamic features. However, as far as I know, such approaches have not been yet used in the macroeconomic multiplier/pass-through literature.

An average time series estimator. There is alternative and simpler approach one can employ to avoid the biases, which works regardless of the dynamic features of the cross section, and is appropriate when T is sufficiently large. The approach involves employing standard time series methods to measure the dynamic effect at horizon h, separately for each i. Thus, one needs to identify policy impulses for each i, and to get rid of the endogenous feedback, the techniques mentioned in the introduction can be used. The estimates obtained will be consistent since the estimated regression coefficients are consistent and dynamic effects are a continuous functions of these coefficients. The cross-sectional mean will then consistently estimate the common component of the dynamic effect at horizon h. This average estimate will be the time series counterpart of β_h in equation (1).

Apart from giving a consistent estimate, such an approach also provides an estimate of the cross-sectional distribution of the effects at each h, which is useful in many applied settings. For example, when considering fiscal (monetary) impulses unit specific estimates could be clustered using economic indicators or type of budget restrictions (banks balance sheet or regulatory information). One could also compute counterfactuals, for example, imposing that the policy impulse has same impact effect across units, but allow the dynamic propagation to be unit specific. None of this information can be obtained with cross-sectional estimates.

Unfortunately, when T is moderate or small, time series estimate of the individual dynamic effects are likely to be biased, and an average of biased estimates is, in general, biased. If the biases happen to be sufficiently idiosyncratic, they will cancel out when computing a cross-sectional average. However, in any relevant policy context, this is unlikely to be the case.

An average estimator for small T. Pooling units in groups which are dynamically homogeneous generally helps when T is small, since the effective sample size becomes $T \times N_m$, where N_m is the size of group m. But, for ungrouped macroeconomic data featuring dynamic heterogeneity, complete pooling is not an option because estimates are more precise but biased.

When T is small, one can exploit the structure of heterogeneity to partially pool units in estimation and thus gain degrees of freedom without introducing obvious biases. The approach is appealing because it works when units are dynamically homogeneous and when they are not, and has a long history in econometrics. One relevant application of the methodology with spatial macroeconomic data is Canova and Pappa [2007].

How does the approach work? Let ϕ_i be the vector of estimated coefficients and assume that the process generating the data is as in (3). Parametrize the heterogeneity as in section 4, i.e. assume that $\phi_i = \phi + v_i$, where v_i a vector of iid random variables, normally distributed, with zero mean and covariance Σ_v. Again, Σ_v measures the extent of dynamic heterogeneity and for $\Sigma_v = 0$ the cross section is dynamic homogeneous. If e_{it} are also normally distributed with zero mean and variance σ_i^2, the DGP and the parametric representation of the heterogeneity imply that ϕ_i is normal with mean $\tilde{\phi}_i = (\frac{1}{\sigma_i^2}x_i'x_i + \Sigma_v^{-1})^{-1}(\frac{1}{\sigma_i^2}x_i'x_i\hat{\phi}_i + \Sigma_v^{-1}\phi)$, where $\hat{\phi}_i$ is the estimator of ϕ_i computed using unit i data. In words, a partial pooling estimator of the parameters is a weighted average of the information contained in unit i data and of the average ϕ assumed, with weights given by the precision of the two types of information. Clearly, if Σ_v is large, that is, when the parameters of the model are very different across units, $\tilde{\phi}_i \to \phi_i$ and an average partial pooling estimator reproduces the average time series estimator discussed above. On the contrary, if Σ_v tends to zero, $\tilde{\phi}_i \to \phi$ and if ϕ when is appropriately chosen, the average partial pooling estimator will come close to a pooled estimator. For any intermediate value of Σ, the average partial pooling estimator will be a linear combination of these two extremes.

The formulas assume that ϕ, Σ_v and σ_i^2 are known or estimable. σ_i^2 can be estimated, unit by unit, on a pre-sample of data. ϕ and Σ_v can be selected in a number of ways; they could reflect pre-sample time series information, spatial information, or aggregate information. For instance, if one has T_1 extra observations, she could employ them to estimates ϕ_i, unit by unit. ϕ will measure the cross sectional mean of ϕ_i and Σ_v the cross sectional dispersion. Alternatively, estimates of ϕ and Σ could be constructed using spatial information. For example, when measuring the effects of fiscal surprises on output and prices in EU countries, for which they had only 16 quarters of data, Canova and Pappa [2007] construct a partial pooling estimator combining each country information with the average effects of fiscal surprises on output and prices for US states. Finally, ϕ and Σ could be estimated with aggregate macroeconomic data. In this case, $\tilde{\alpha}_i$ combines unit specific and aggregate information using the relative precision of the two types of information as weights.

6 Some experimental evidence

To quantitative measure the relative distortions that various estimators produce, I run a small Monte Carlo exercise, using N=50, T=40, and Q=1000 replications. The data for each

$i = 1, \ldots, N$ is generated by:

$$y_{1,i,t} = a_{1,i} + b_{1,i}y_{1,i,t-1} + b_{2,i}y_{2,i,t-1} + d_{1,i}x_t + e_{1,i,t} + 0.5e_{2,i,t} + c_{1,i}e_{3,i,t} \quad (7)$$

$$y_{2,i,t} = a_{2,i} + b_{2,i}y_{1,i,t-1} + c_{2,i}v_{i,t} + d_{2,i}x_t + e_{2,i,t} + 0.5e_{1,i,t} + c_{2,i}e_{3,i,t} \quad (8)$$

$$p_{i,t} = a_{3,i} + b_{3,i}p_{i,t-1} + c_{3,i}e_{1,i,t} + e_{3,i,t} \quad (9)$$

$$x_t = 0.95\ x_{t-1} + 0.1\ u_{i,t} \quad (10)$$

$$z_{i,t} = b_{4,i}p_{i,t} + v_{i,t} \quad (11)$$

where $e_{i,t}, u_{i,t}, v_{i,t}$ are assumed to be iid with zero mean, with unit variance, and zero covariance. Thus, the DGP features two interdependent autoregressive endogenous variables $y_{j,i,t}, j = 1, 2$, driven by an exogenous autoregressive variable x_t, by the policy variable $p_{i,t}$, and by two variable specific disturbances. The policy variable is endogenous and responds to the innovations in $y_{1,i,t}$; the instrument $z_{i,t}$ is randomly related to the policy variable.

The vector of parameters $\phi_i = (a_i = (a_{1,i}, a_{2,t}, a_{3,t}), b_i = (b_{1,i}, b_{2,i}, b_{3,i}, b_{4,i}), d_i = (d_{1,i}, d_{2,i}), c_i = (c_{1,i}, c_{2,i}, c_{3,i})$ is assumed to be normally distributed. In the baseline specification the mean is $\phi_0 = ((0, 0, 0), (0.6, 0.3, 0.9, 0.2), (1.6, 0.05), (1.0, 0.8, 0.7))$ and the covariance matrix is $\omega * I$, where $\omega = (\omega_1, \omega_2)$. ω_1 applies to all the components of ϕ_i except to $c_{1,i}$ and $d_{1,i}$ for which ω_2 is used instead. ω dictates the amount of dynamic heterogeneity: for $\omega_1 = \omega_2 = 0$ the cross-section is dynamically homogeneous; for $\omega_1 = 0, \omega_2 = 2.5$ the cross section is instantaneous heterogeneous but the lag dynamics are homogeneous; for $\omega_1 = 0.1, \omega_2 = 0$ it is instantaneously homogeneous but lagged dynamics are heterogeneous; and for $\omega_1 = 0.1, \omega_2 = 1.5$, the cross-section displays both instantaneous and lagged heterogeneity. b_1, b_3 are selected to match the persistence of output and government spending in the sample of US states used in the next section, while c_1 is chosen so that the cross-sectional distribution of instantaneous output (which I take to be $y_{1,t}$) response has a mean of about 0.8.

I am also interested in examining the robustness of the conclusions when the persistence of the autoregressive process for $y_{1,t}$ is reduced (the mean of $b_{1,i}$ drops from 0.6 to 0.2); the correlation between the policy variable and the instrument is increased (the mean of $b_{4,i}$ increases from 0.2 to 0.8); cross variable feedbacks are reduced (the mean of $b_{2,i}$ drops from 0.3 to 0.15); and the policy variable is less endogenous (the mean of $c_{3,i}$ drops from 0.7 to 0.1).

I assume that the data is annual and I report results for the two-year horizon. I estimate equation (1) in a number of ways. In the first (model A), there are fixed and time effects and the equation is estimated by cross-sectional IV methods with $z_{i,t}$ as instrument; in the second (model B), there are fixed and time effects, cross-sectional IV methods are employed but the policy variable enters in first difference ($p_{i,t} - p_{i,t-1}$) and the instrument is also differenced $((z_{i,t} - z_{i,t-1}))$; in the third (model C) equation (1) is estimated with cross-sectional methods allowing for fixed effects, one lag of the dependent variable, while the policy variable and instruments remain in first difference; in the fourth (model D) the specification and estimation are as in model C, but the instruments are lagged enough to insure that they are uncorrelated with the composite error term.

I also estimate (4), unit by unit (Model E), compute unit specific effects using $\hat{m}_i = \hat{c}_i \sum_{h=1}^{2} \hat{b}_i^h$ and take the cross-sectional average $\hat{m} = \frac{1}{N}\sum \hat{m}_i$ as an estimate of the mean effect. I also generate data for $N_1 = 20$ additional units, estimate model (4) with this data, unit by unit, obtain a cross-sectional mean and the dispersion matrix and use them to construct a partial pooling estimator (Model F). For models E and F, an IV approach is also used, with $z_{i,t}$ as instrument. I report the mean square error (MSE) computed as the difference between the estimate and the true effect, on average across replications. Other summary statistics, such as the mean absolute deviation, give similar conclusions.

Results for the baseline cases. Table 1 presents the results. With instantaneous and lagged heterogeneity (row 1), all four cross-sectional estimators fail and the MSEs are 50 percent larger than with the alternative estimators. Because T is short, using additional information helps in regularising average estimates. Given the dispersion of MSE estimates, one can not reject the hypothesis that the partial pooling estimator is unbiased.

When only instantaneous heterogeneity is present (row 2), the average partial pooling estimator is the best and the average time series estimator is also superior to all cross-sectional estimators. When only lagged heterogeneity is present (row 3), the ranking is reversed: the average time series estimator is superior to the average partial pooling estimator. The four cross-sectional estimators are roughly equivalent in MSE terms and lag behind. Finally, when the cross section is dynamically homogeneous (row 4), the estimator obtained with model C is best. Thus, when the cross-section is dynamically homogeneous, properly accounting for general equilibrium interrelationship seems preferable to employing an all-purpose time

Table 1: Two years ahead MSE						
DGP	Model A	Model B	Model C	Model D	Model E	Model F
Baseline exercises						
$\omega_1 = 0.1, \omega_2 = 1.5$	0.3418	0.3605	0.3359	0.3388	0.2303	0.2299
	(0.1351)	(0.3988)	(0.1279)	(0.1277)	(0.1325)	(0.1199)
$\omega_1 = 0.0, \omega_2 = 2.5$	0.1816	0.4067	0.2141	0.8879	0.18032	0.1467
	(0.1183)	(0.2746)	(0.1842)	(1.2788)	(0.2242)	(0.1397)
$\omega_1 = 0.1, \omega_2 = 0.0$	0.33472	0.34773	0.33323	0.33552	0.28133	0.28491
	(0.0392)	(0.3325)	(0.0404)	(0.0349)	(0.0507)	(0.0432)
$\omega_1 = 0.0, \omega_2 = 0.0$	0.1654	0.3591	0.1191	0.7000	0.2343	0.2366
	(0.0304)	(0.0758)	(0.0472)	(0.6444)	(0.0357)	(0.0249)
Additional exercises						
$\omega_1 = 0.1, \omega_2 = 1.5$	0.0278	0.0278	0.0278	0.0280	0.0155	0.0113
$b_{10} = 0.2$	(0.0071)	(0.0073)	(0.0073)	(0.0073)	(0.0226)	(0.0140)
$\omega_1 = 0.0, \omega_2 = 0.0$	0.0063	0.6056	0.4593	1.0795	0.0113	0.0055
$b_{10} = 0.2$	(0.0049)	(0.1371)	(0.1274)	(1.1427)	(0.0197)	(0.0092)
$\omega_1 = 0.1, \omega_2 = 1.5$	0.3418	0.3564	0.3358	0.3386	0.2320	0.2201
$b_{40} = 0.8$	(0.1350)	(0.3396)	(0.1282)	(0.1277)	(0.2015)	(0.1495)
$\omega_1 = 0.0, \omega_2 = 0.0$	0.1638	0.3619	0.1763	0.7059	0.2170	0.2161
$b_{40} = 0.8$	(0.0164)	(0.0284)	(0.0180)	(0.2070)	(0.0990)	(0.0524)
$\omega 1 = 0.1, \omega_2 = 1.5$	0.2633	0.2625	0.2618	0.2640	0.1948	0.1967
$b_{20} = 0.15$	(0.1113)	(0.1121)	(0.1118)	(0.1114)	(0.1087)	(0.0996)
$\omega_1 = 0.0, \omega_2 = 0.0$	0.1329	0.3801	0.1546	0.7063	0.1769	0.1792
$b_{20} = 0.15$	(0.0262)	(0.0801)	(0.0567)	(0.6720)	(0.0305)	(0.0206)
$\omega 1 = 0.1, \omega_2 = 1.5$	0.3423	0.3958	0.3365	0.3391	0.2360	0.2361
$b_{30} = 0.1$	(0.1375)	(1.0938)	(0.1274)	(0.1281)	(0.1470)	(0.1219)
$\omega_1 = 0.0, \omega_2 = 0.0$	0.1695	0.3292	0.0934	0.8648	0.2381	0.2411
$b_{30} = 0.1$	(0.0435)	(0.1613)	(0.0767)	(1.0589)	(0.0403)	(0.0310)

Notes: In parenthesis is the dispersion of MSE estimates across replications. In the first row the DGP displays instantaneous and lagged heterogeneity; the second only instantaneous heterogeneity; the third only lagged heterogeneity; in the fourth row the DGP is dynamic homogeneous. The fifth and sixth rows report results when the DGP is fully heterogeneous or homogenous cases and low persistence in the process for $y_{1,i,t}$; the seventh and eight rows results when stronger instruments are used; the ninth and tenth rows results when there is less interdependences in the data; and the last two rows results when the policy variable is more exogenous.

effects, see also Ramey [2020]. With dynamic homogeneity, the two alternative estimators perform well and are superior to two cross-sectional estimators.

Two additional points regarding the performance of cross-sectional estimators are useful. First, model B estimator is never the best of the four cross-sectional estimators and displays a lot of cross-replication dispersion. Its poor performance is due to the fact that, with time effects and the policy variable in difference, misspecification is important (the policy variable is over-differenced). Second, model D always produce worse MSE estimates than model C. Because the only difference between the two specifications is the timing of the instruments, there is an important trade-off between making the instruments more exogenous and reducing their relevance for the policy variable.

Robustness The ranking of estimators does not change significantly with the specification of the design. When there is dynamic heterogeneity, proceeding unit by unit and averaging is preferable, and this is true regardless of whether instantaneous or lagged heterogeneity, or both are present. When instead the cross section is dynamic homogeneous, a standard specification with time and fixed effects is the least distorting. Quantitatively, the magnitudes of the MSE reported in the first part of the table are also not substantially affected by changes in the parameters of the design.

7 Local fiscal multipliers in US states

To show that the differences found in the experimental exercise are also important in practical situations and lead to opposite economic conclusions, I employ US state data to construct a measure of the average local government expenditure multiplier. The data is annual and goes from 1980 to 2017. Because the BEA currently publishes only data from 1993, I splice the data with the one used by Canova and Pappa [2007] to have a longer time series. I compute one- and two-years multipliers. The same patterns I present also hold when $h = 3$ or 4.

The alternative specifications. The first specification I estimate is similar to Chodorow-Reich [2019]. The cumulative difference in the real gross state product between $t + h$ and t is regressed on fixed and time effects, and the level of state government expenditure at t. No additional controls are included. Given the endogeneity of government expenditure, cross-sectional IV estimation is performed, using contemporaneous federal government transfers as instrument (Model A). The second specification is similar Nakamura and Steinnson [2014]. Here the percentage growth rate of the real state product between $t+h$ and t is regressed on fixed and time effects, and the cumulative change in state government expenditure, scaled by gross state product at t. I use a cross-sectional approach and instrument the cumulative change in the scaled expenditure variable with the cumulative change in scaled federal government transfers (MODEL B). The third specification adds to model B a lag of the dependent variable. Estimation is by cross-sectional IV methods (MODEL C). The fourth specification omits the time effect and uses the lagged values of state taxes, the current oil price growth, and the US real rate of interest as controls. Estimation is also conducted with cross-sectional IV techniques using one lag of the cumulative change in scaled federal government transfers as instrument. At the two years horizon, the cumulative change in the scaled expenditure variable is instrumented with the cumulative change in scaled federal transfers lagged two periods, to eliminate the correlation between the instruments and the composite error term (MODEL D). Since models A-D include either a time effect or aggregate variables supposed to capture the US cycle, the coefficients on the instrumented expenditure variable are estimates of the average cross-sectional multipliers.

I consider two additional specifications. In both cases, the regressors are the same as in model D. In Model E, I estimate the specification unit by unit for each h, compute a cross-sectional mean and a measure of dispersion (MODEL E). In model F, I first estimate model D using aggregated US data. I then assume that the coefficients of the state regressions θ_i have mean equal to $\theta_{US,h}$ and covariance matrix $\lambda_h * cov(\theta_{US,h})$, where λ_h is a scalar controlling cross-sectional heterogeneity. I set λ_h as a function of the horizon ($\lambda_1 = 0.02, \lambda_2 = 0.003$) to account for the fact that, with short data, estimates at longer horizons are likely to be poorer.

The results. Table 2 reports the results. Average cross-sectional multipliers are generally negative and their standard errors are large. Thus, with most specifications, one cannot reject the hypothesis that they are zero. The exception is model A. However, since the policy regressor is the level of expenditure rather than its innovation (or a proxy for it), an upward bias in the estimate is likely to be present, given that local spending is positively serially correlated. Specifications which include macroeconomic variables rather than time effects and properly lagged instruments, produce point estimates of the average multipliers which are generally larger in absolute value. Thus, having controls rather than time effects and proper instruments matters.

Table 2: Average US state government expenditure multipliers

Horizon	Model A	Model B	Model C	Model D	Model E	Model F
1 year	0.079	-0.282	-0.251	-0.498	0.588	0.920
	(0.004)	(0.295)	(0.286)	(0.295)	(0.364)	(0.130)
2 years	0.223	-0.284	-0.097	-1.487	0.965	1.0207
	(0.012)	(0.228)	(0.186)	(1.898)	(1.760)	(0.077)

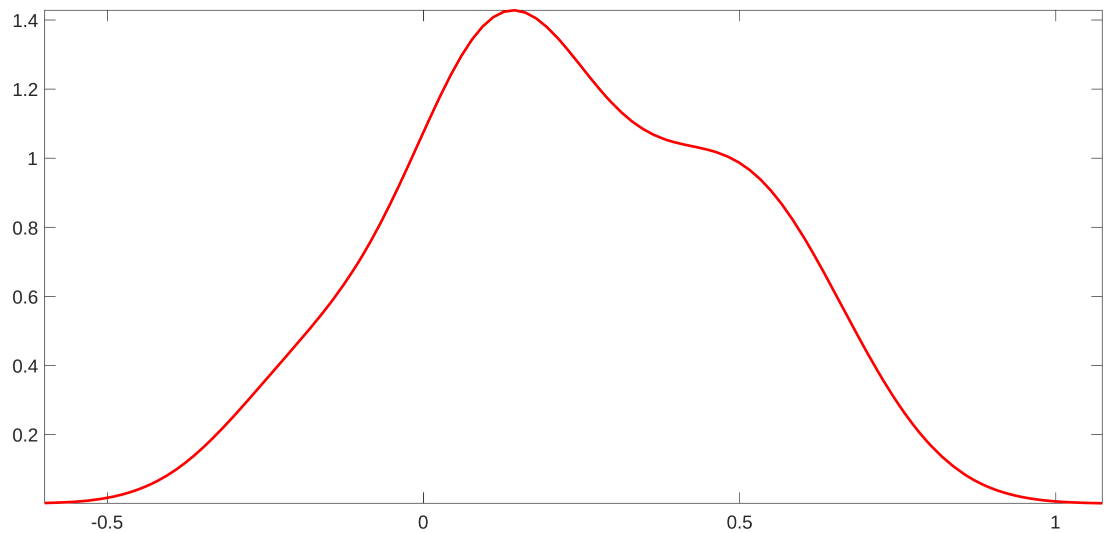
Notes: In parenthesis are the standard errors of the estimates.

Why are cross-sectional multipliers generally negative? One possibility is that dynamic heterogeneity matters. Figure 1, which plots the cross-sectional distribution of estimates of the AR(1) coefficient of gross state output, shows that, indeed, dynamic heterogeneity is important. Thus, none of the cross-sectional average estimate should trusted.

The average multipliers obtained with Models E and F are positive. However, because of the considerable heterogeneity, there is a large dispersion of individual estimates in model E and local government spending multipliers exceeding 2 or falling below -1 are possible. Because the sample is short, partial pooling may help to regularize both the distribution of single unit multiplier estimates and its average. The last column of table 1 shows that, indeed, the information helps: multiplier estimates at both horizons are larger and insignificantly different from one.

The policy conclusions one draws depend on the methodology employed. While cross-sectional estimates indicate that the average local spending multiplier is zero (if not negative), a partially pooled estimate suggests that it is around one. Thus, even though local fiscal policy is unlikely to have large private sector effects, it may help to stabilize gross state product, when needed.

Figure 1: Distribution of gross state output persistence parameter



8 Conclusions

The recent practice of computing fiscal multipliers or monetary policy pass-through with a cross-sectional methodology is justified only under stringent statistical requirements and dynamic homogeneity in the evolution of the cross-section. When dynamic heterogeneity is present, the estimates obtained are invalid, even when IV is employed.

With dynamic heterogeneity, and T sufficiently large, one could estimate the effect of a policy change in time, unit by unit, and then compute a cross-sectional average. When T is short, one can regularize the average estimate by combining unit specific and extraneous information, coming from time series or spatial information, or aggregate data, using partial pooling techniques.

When dynamic homogeneity is questionable, we recommend researchers to estimate the model using time series variations, unit by unit, and plot distribution of the AR(1) coefficient and of the instantaneous policy impact, both of which provide a sense of the importance of dynamic heterogeneity in the data employed.

Spatial analyses can inform macroeconomists about the quality of their models and the effects of government policies, adding a new dimension to standard time series analyses. A cross-sectional methodology to compute macroeconomic objects is nowadays popular, but by no means, a free lunch. Understanding the trade-offs and minimizing the chance of incorrect conclusions is important to make empirical research on monetary and fiscal policies trustworthy.

References

Carlo Altavilla, Fabio Canova, and Matteo Ciccarelli. Mending the broken link: heterogeneous bank lending and unconvetional monetary policy. *Journal of Monetary Economics*, 110:81–98, 2020.

Fabio Canova. *Methods for Applied Macroeconomic Research*. Princeton University Press, 2007.

Fabio Canova and Paraskevi Pappa. Price differentials in monetary unions: the role of fiscal shocks. *Economic Journal*, 117:717–739, 2007.

Indraneel Chakraborty, Itay Goldstein, and Andrew MacKinlay. Monetary stimolous and bank lending. *Journal of Financial Economics*, 136(1):189–218, 2020.

Gabriel Chodorow-Reich. Geographic cross-sectional fiscal spending multipliers: What have we learned? *American Economic Journal: Economic Policy*, 11(2):1–34, 2019.

Bill Dupour and Rodrigo Guerrero. Local and aggregate fiscal policy multipliers. *Journal of Monetary Economics*, 92:16–30, 2017.

Mark Gertler and Peter Karadi. Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76, 2015.

Gabriel Jimenez, Steven Ongena, Jose Luis Peydro', and Jesus Saurina. Hazardous times for monetary policy: What do 23 million loans tell us about the impact of monetary policy on credit risk taking? *Econometrica*, 82:463–505, 2014.

Soyoung Kim and Nouriel Roubini. Exchange rate anomalies in industrialized countries: A solution with a structural var. *Journal of Monetary Economics*, 45:561–583, 2000.

Emi Nakamura and Jon Steinnson. Fiscal stimulus in a monetary union: Evidence from us regions. *American Economic Review*, 103(4):753–792, 2014.

Valerie Ramey. Ten years after the financial crisis: What have we learned from the renaissance in fiscal research. *Journal of Economic Perspectives*, 33:89–114, 2019.

Valerie Ramey. Discussion of 'what do we learn from cross-sectional empirical estimates in macroeconomics' by adam guren, alisdair mckay, emi nakamura, jon steinsson. Technical report, University of California San Diego, 2020.

Valerie Ramey and Sarah Zubairy. Goverment spending multipliers in good times and in bad: Evidence from us historical data. *Journal of Political Economy*, 126:850–901, 2018.