

Identification, estimation and inference in Panel Vector Autoregressions using external instruments

Raimondo Pala*

November 25, 2025

Abstract

This paper proposes an identification inspired from the SVAR-IV literature that uses external instruments to identify PVARs, and discusses associated issues of identification, estimation, and inference. I introduce a form of local average treatment effect - the μ -LATE - which arises when a continuous instrument targets a binary treatment. Under standard assumptions of independence, exclusion, and monotonicity, I show that externally instrumented PVARs estimate the μ -LATE. Monte Carlo simulations illustrate that confidence sets based on the Anderson-Rubin statistics deliver reliable convergence for impulse responses. As an application, I instrument state-level military spending with the state's share of national spending to estimate the dynamic fiscal multiplier. I find multipliers above unity, with effects concentrated in the contemporaneous year and persisting into the following year.

JEL Classification: C32, C33.

Keywords: Vector autoregressive models, panel data, identification, instruments.

*Email: raimondopala@gmail.com

Panel vector autoregressions (PVAR) are one of the standard tools for the estimation of dynamic causal effects in macroeconomics. However, their causal interpretation is frequently limited. Sometimes they are described as identifying a temporally related causality, “Granger causality”; sometimes, as discussed in [Pala \(2025\)](#), they may posses a contemporaneous causal interpretation under either exogeneity assumptions, or a suitable control type of assumption. However, Granger causality is not contemporaneous causality¹; exogeneity is hard to satisfy due to the rich endogeneity among macroeconomic variables ([Nakamura and Steinsson \(2018\)](#)); and the control type of approach proposed by [Pala \(2025\)](#) may not be satisfied if some units cannot act as controls for some others. However, when all the previous cases fail, instruments could be utilised to carry causal claims.

This paper shows that it is possible to identify causal effects by the means of instrumental variables in the case of panel vector autoregressions. By implementing the same approach as [Gertler and Karadi \(2015\)](#); [Mertens and Ravn \(2013, 2014\)](#); [Stock and Watson \(2018\)](#); [Olea et al. \(2021\)](#) - defined as Proxy-SVAR or SVAR-IV- in PVARs it is possible to retrieve a causal estimand that I define as μ – LATE. Normally, the causal literature has a clear target in mind in the case of a dummy policy and a dummy instrument ([Angrist et al. \(1996\)](#)) - the LATE. LATEs can be viewed as a special case of principal stratification for which there are 4 categories of compliance status. In the case of continuous instruments the interpretation becomes troublesome because there are infinitely many possible strata ([Antonelli et al. \(2023\)](#)). Contrarily, μ -LATEs seemlessly move from continuous variables to a binary interpretation akin to a LATE. In particular, they compare the value of the outcome and policy under a 1% and 0% instrument assignment.

In this special context, the defiers are (under positive monotonicity) units that do not observe an increase (decrease) in their policy variable residuals when the instrument increases (decreases). Moreover, because usually the residuals of a PVAR are assumed to be continuous and normally distributed, the μ - LATE strongly depends on the instrument

¹In the words of [Granger and Newbold \(2014\)](#): “A better term might be temporally related, but since it is such a simple term we shall continue to use it .

and policy continuity. This latter feature is rarely discussed by the causal literature but imposes a strong linearity assumption.² This is, however, necessary. Because macroeconomic data is frequently relatively small in the unit component, making use of a linearity assumption to obtain meaningful estimates is often the only possibility.³

On the other hand, inference can be carried by simply translating the conclusions of Olea et al. (2021) regarding VAR-IVs to their panel counterpart. This implies that the optimal approach to inference is to make use of the Anderson Rubin test statistic (henceforth defined as AR - see Fieller (1944); Anderson and Rubin (1949) for its definition) in the presence of one instrument, and, at the current state of the literature, the conditional likelihood ratio test (henceforth defined as CLR - see Andrews et al. (2006) for its definition) in the presence of multiple instruments.

I give a context to the proposed causal framework through an original application in which I compute the dynamic fiscal military multiplier in the United States. Using local level data can be advantageous for several reasons. First, Nakamura and Steinsson (2014) argue that in a monetary union, the central bank (the Federal Reserve) cannot raise interest rates in some states relative to others, and federal tax policy is common across states in the union. This means that the open economy multiplier identifies the measure of fiscal policy that does not depend on monetary movements. Second, the advantage of using military spending is that the DD-350 military procurement forms made available from the US Department of Defense provide a clean, direct, and localised measure of government spending. Unfortunately, data of such quality, that starts from such an early age - 1966 - is not available for any other form of spending.

For this reason, I consider a PVAR that includes military state-level spending as instrumented by the states fraction of total national military spending, with GDP growth as the outcome variable. Such IV approach is generally defined as Bartik instrument and is becoming increasingly popular in social sciences, where the use of local-level data is

²In fact, most of the IV literature has focused on either binary (Angrist et al. (1996)) or multi-valued instruments (Heckman and Vytlacil (2001); Vytlacil (2002)), but rarely discussed how a LATE could emerge from the comparison of two predicted values of the outcome variables under a continuous instrument case. In general, semi parametric or non parametric estimators tend to be preferred when the data allows it.

³This point, and its many drawbacks, are developed in Kolesár and Plagborg-Møller (2024).

intensifying (see [Bartik \(1991\)](#); [Goldsmith-Pinkham et al. \(2020\)](#)). My findings indicate that the fiscal multiplier may be about ~ 1.7 on impact, a value not dissimilar from the ones previously observed by the state-multiplier literature ([Chodorow-Reich \(2019\)](#)). The advantage of using PVARs over simple panel linear regressions becomes clear when considering the usefulness of Impulse Response Functions. In this sense, my findings suggest a multiplier of about 1.5 one year ahead and negligible after, suggesting the existence of large cumulative multipliers.

This paper contributes to several streams of the literature. One natural contribution is at the intersection of the causal inference and the time series fields. In this sense, it complements [Pala \(2025\)](#) in the research agenda oriented to give a causal interpretation to panel vector autoregressions and, more broadly, the literature that uses the Rubin causal model to motivate the causal interpretation of time series models ([Menchetti and Bojinov \(2022\)](#); [Menchetti et al. \(2022\)](#); [Rambachan and Shephard \(2021\)](#); [Bojinov and Shephard \(2019\)](#)). Second, it contributes to the econometric literature by extending the identification of structural VARs through an external instrument to panel VARs. Third, it indicates a potentially interesting recipe for the interpretation of any estimator in the case of continuous instruments that is different from the current literature. Fourth, it extends the common knowledge of the good coverage properties of the Anderson-Rubin statistics for the impulse response functions generated by a PVAR-IV, indicating a venue for the construction of reliable confidence sets. Fifth, it contributes to the applied literature for the estimation of the dynamic fiscal multipliers.

The discussion is organized as follows. Section 1 introduces the potential outcome framework and the μ – LATE. Section 2 introduces the assumptions required for the identification of the causal effects. Section 3 introduces the estimation procedure of the PSVAR-IV. Section 4 discusses the Anderson Rubin statistics⁴. Section 5 discusses the estimation of the dynamic fiscal multiplier using the PSVAR-IV.

Finally, allow me to introduce some useful notation. A set of observations of variable s for state i at time t , $x_{s,it}$, is part of a larger matrix of variables \mathbf{x}_s , and $\mathbf{x}_{2:S}$ indicates

⁴In appendix 7.3 I provide the main theorems and simulations that showcase the good coverage properties of the statistic.

the partition from the second to the S variable of such matrix. $\mathbb{E}(X)$ will indicate the expected value of the variable X , $var(X)$ will indicate its variance, $cov(X, Y)$ the covariance between variables X and Y , and $cov(X, Y)$ their variance-covariance matrix, which contains their variance on the diagonal and their covariance in the off diagonal. For R being a square matrix, $R_{a,b}$ indicates the entry of R in the row a and column b . $X \perp Y$ will indicate independence between X and Y .

1 Identification

Let us say the researcher is interested in the estimation of a dynamic causal effect, represented as the impact of a change in a *policy* variable $W_{i,t}$ on one or many *outcome* variables $Y_{j,i,t}$. As such, the potential outcomes of the outcome variables can be represented as follows:

$$Y_{j,i,t}(w, z) = Y_{j,i,t}((W_{i,1:t-1}, w, W_{i,t+1:T})(Z_{i,1:t-1}, z, Z_{i,t+1:T})).$$

Such definition is similar to the one introduced by [Rambachan and Shephard \(2021\)](#) and indicates that the potential outcome of the different outcome variables j depend on the policy assignment w and on the instrumental variable z ⁵. In the leading example of this paper, there will only be GDP growth as outcome variable ($j = 1$), while $W_{i,t}$ will indicate a fiscal policy variable. Moreover, Z_{it} will be the instrumental variable. Finally, the index i, t will refer to region i at time t .⁶ All variables are assumed to be continuous.

I define a μ -LATE $_j$ as the following causal effect.

Definition 1. A μ -LATE $_j$ is the LATE of receiving assignment w versus assignment w' for those units that comply with the assignment

$$\mu\text{-LATE}_j = \mathbb{E}\left[Y_j(w) - Y_j(w') \mid W(z) = w\right].$$

⁵Such form was first introduced in LATEs by [Angrist et al. \(1996\)](#).

⁶In this paper, I will solely focus on a clean case of one instrument, one policy. This choice stems from the higher likelihood of researchers to come up with identification strategies that involve a clear one instrument, one instrumented, many outcomes scenarios. It can be extended to multiple instruments and multiple instrumented variables. The analytical extensions are provided in [Olea et al. \(2021\)](#).

2 Assumptions

The assumptions required to give a causal interpretation to a LATE estimator are independence, exclusion and monotonicity. Here the assumptions are put on the residuals of the PVAR. In particular, define $\tilde{W}_{it} = \mathbb{E}[W_{it}|\omega_{it}]$ and $\tilde{Y}_{j,it} = \mathbb{E}[Y_{j,it}|\omega_{it}]$, where $\omega_{it} = (W_{i,t-1}'..W'_{i,t-p}, Y'_{1,i,t-1},.., Y'_{1,i,t-p}, Y'_{J,i,t-1},.., Y_{J,i,t-p})$ are the regressors. The PVAR estimates the impact of variations in \tilde{W}_{it} on $\tilde{Y}_{j,it}$.

The treatments are assumed to be continuous, so that \tilde{w}° is any value extracted from the distribution of \tilde{W} and z° is any value extracted from the distribution of Z .

Assumption 1. (Independence) For all $\tilde{w}^\circ \in \tilde{W}$, all $\tilde{z}^\circ \in Z$, all $t \geq 1$, all $i \geq 1$, and all $j \geq 1$, it holds that

$$\{\tilde{Y}_{j,it}(\tilde{w}^\circ, z^\circ), \tilde{W}_{i,t}(z^\circ)\} \perp Z_{i,t} \quad (1)$$

Assumption 1 establishes that the instruments are as good as randomly assigned with respect to any of the outcome variable's residuals or any of the policy variable's residuals. The researcher will normally assume that the potential outcomes are not affected by Z_{it} if not through \tilde{W}_{it} , i.e.

Assumption 2. (Exclusion) For all $\tilde{w}, \tilde{w}' \in \tilde{W}_{i,t}$, $t \geq 1$, and $i \geq 1$ it holds that

$$\{\tilde{Y}_{j,it}(\tilde{w}, z) = Y_{j,it}(\tilde{w}, z')\}$$

where $z, z' \in Z_{i,t}$ are any possible combination of values of $Z_{i,t}$.

Notice that one of the consequences of assumption 2 is that it implies that the potential outcomes of $\tilde{Y}_{j,it}$ - the residual of the outcome variable - do not depend on the realized value of the instruments, if not by the means of the policy. Such assumption, coupled with assumption 1, allows for a seamless transition from conditional expected values and realized outcomes to potential outcomes frameworks. In the empirical example, this would mean that the assignment of military expenditures at the federal-level are independent

with respect to the potential outcome process of any GDP growth innovations and with respect to any military spending innovations at the state-level.

Finally, monotonicity as in Angrist et al. (1996) is required, so that

Assumption 3. (Monotonicity) For all $z, z' \in Z_{i,t}$, and all $t \geq 1$ and $i \geq 1$, it holds that either $\widetilde{W}_{i,t}(z) \geq \widetilde{W}_{i,t}(z')$ or $\widetilde{W}_{i,t}(z) \leq \widetilde{W}_{i,t}(z')$.

Notice that this type of monotonicity assumption needs to hold for every couple of instrument assignments z, z' . For example, in the case of fiscal multipliers, it means that if there is an increase in total national military spending, the residuals of state-level military spending are either increasing or decreasing for each unit i at each time t . Such assumption, differently from a non parametric case, does not allow any discontinuity of the mapping of \widetilde{W} on Z .

Remark 1. (Weakness of a fully parametric monotonicity assumption). Let us say that Z is multivalued, such that it can only take three integer values $Z = \{z^0, z^1, z^2\}$. A parametric estimator needs to assume $W(z^2) \geq W(z^1) \geq W(z^0)$. A non parametric estimator could potentially solve this issue by estimating two separate quantities, one for z^1 and z^2 , and assuming that $W(z^1) \geq W(z^0)$, $W(z^2) \geq W(z^0)$, but the ordering of $W(z^2) \underline{W}(z^1)$ does not need to be assumed. This comes at the cost of requiring the data to be dense enough around the quantities.

Finally, the instrument needs to be a predictor of the instrumented variable, a condition frequently defined as *instrument relevance* or *instrument strength*.

Assumption 4. (Relevance). The instrument satisfies $\mathbb{E}[\widetilde{W}_{i,t}, Z_{i,t}] \neq 0$.

Remark 2. Frequently, SVAR-IV are thought to estimate a causal effect only under a relevance and an non correlation condition. In macro economics, the non correlation condition is frequently stated as $\mathbb{E}[\widetilde{Y}_{j,it}, \widetilde{W}_{i,t}] = 0$ ⁷. Yet, such imposition would only refer to a statistical relationship among the two, and would ignore the benefits of having a

⁷See, among many others, Stock and Watson (2018); Olea et al. (2021); Bruns and Keweloh (2024); Brignone et al. (2023).

potential outcome representation⁸. In practice, this means that SVAR-IV would not be able to claim that a meaningful causal estimand has been identified unless the econometrician identifies a set of assumptions that maps the estimator to an estimand. Typically, the two conditions that are sought for are insufficient to achieve any causal identification.

3 Estimation

Consider several known outcome variables and one intervention variable aggregated in a process of the kind

$$x_{i,t} = (W'_{i,t}, Y'_{j=1,i,t}, Y'_{j=2,i,t}, \dots, Y'_{j=J,i,t})'$$

Here $W_{i,t}$ could indicate military procurement spending in region i at time t , and $j = 1, \dots, J$ could indicate output and other outcomes of interest in region i at time t . PVARs are generally represented as processes that depend on their past, a series of unit-specific characteristics, and some random disturbances, which leads to:

$$x_{i,t} = (I_m - \Phi)\mu_i + \Phi x_{i,t-1} + \tilde{x}_{i,t} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (2)$$

Here Φ denotes an $m \times m$ matrix of slope coefficients⁹, μ_i is an $m \times 1$ vector of individual-specific effects, $\tilde{x}_{i,t}$ is an $m \times 1$ vector of disturbances, and I_m denotes the identity matrix of dimension $m \times m$. The model can be extended to include higher lags, but to keep the notation compact I will use a one lag representation.

The focus of the following section will be on the disturbances

$$\tilde{x}_{i,t} = (\widetilde{W}'_{i,t}, \widetilde{Y}'_{j=1,i,t}, \widetilde{Y}'_{j=2,i,t}, \dots, \widetilde{Y}'_{j=J,i,t})',$$

where the tilde represents the specific disturbance related to the original variable. Such

⁸In this sense, the condition usually stated is potentially testable (see [Bruns and Keweloh \(2024\)](#)), but does not allow to claim the identification of a LATE.

⁹This framework could be extended to random effects instead of fixed effects. However, such change would have no impact on the nature of the causal effects estimated.

disturbances are distributed according to $\tilde{x}_{i,t} \sim \mathcal{N}(\mathbf{0}_{J+1}, \Sigma)$ where \mathcal{N} is a normal distribution. In the rest of the paper will assume that the policy variable goes first¹⁰, and the outcome variables follow. For example, in the case of fiscal multipliers this would mean that $\tilde{W}_{i,t}$ refers to the innovations of military spending growth in each region of the US, and $\tilde{Y}_{1,i,t}$ refers to the innovations of GDP growth in those same regions. Exactly like vector autoregressions, panel vector autoregressions have a contemporaneous causal representation that is commonly defined as panel structural vector autoregression (PSVAR), as follows

$$Rx_{i,t} = R(I_m - \Phi)\mu_i + R\Phi x_{i,t-1} + R\tilde{x}_{i,t} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (3)$$

Generally, the estimation of the contemporaneous causal effects is carried on transformations of $\Sigma = R^{-1}R^{-1'}$, where R^{-1} is the unique lower-triangular Cholesky factor with non-negative diagonal elements. The reduced-form innovations \tilde{x}_{it} are related to the SPVAR shocks η_t by an invertible matrix H :

$$\tilde{x}_{i,t} = H\Gamma\eta_{i,t} = R^{-1}\eta_{i,t}, \quad \eta_{i,t} \sim (0, I_{J+1}), \quad \text{diag}(H) = 1,$$

where $R^{-1} = H\Gamma$, and Γ is a diagonal matrix with variance of the shocks in the diagonal entries. The structural shocks $\eta_{i,t}$ are mean zero with unit variance, serially and mutually uncorrelated. Since the autoregressive parameters $\hat{\Phi}$ can be consistently estimated under regularity conditions, the sample residuals $\hat{\tilde{x}}_{i,t}$ are consistent estimates of $\tilde{x}_{i,t}$. The empirical SPVAR problem reduces to finding R from $\hat{\Phi}$. But there are $(J+1)^2$ parameters in R and the sample covariance of $\hat{\tilde{x}}_{i,t}$ only provides $(J+1)((J+1)+1)$ conditions in face of $(J+1)^2$ parameters to be estimated. The SPVAR is therefore under-identified as there can be infinitely many solutions that satisfy the covariance restrictions.

The IV procedure for the estimation of the structural matrix in SVARs generally corresponds to either one or two separate steps, depending on whether the economist is interested in a unit normalized shock or a standard shock¹¹. First, an IV is estimated

¹⁰This is a standard assumption in the SVAR-IV literature (see [Stock and Watson \(2018\)](#)) .

¹¹See [Gertler and Karadi \(2015\)](#); [Mertens and Ravn \(2013, 2014\)](#); [Stock and Watson \(2018\)](#); [Olea](#)

with the following first and second stages

$$\widetilde{W}_{i,t} = \delta Z_{i,t} + \eta_{i,t} \text{ first stage}$$

$$\widetilde{Y}_{j,i,t} = \beta_j \widetilde{W}_{i,t} + \epsilon_{i,t} \text{ second stage(s)}$$

Then, if the economist is interested in a unit normalized shock, the IV estimator simply becomes $\beta_j^{IV} = (\rho_j/\delta)$, where ρ_j comes from the regression $\widetilde{Y}_{j,i,t} = \rho_j Z_{i,t} + \nu_{i,t}$, $R_{1,1} = \delta$, and $R_{1,j} = \beta_j^{IV}$. In the case in which the economist is interested in a standardized shock, instead, the IV estimator becomes $\beta_j^{IV} = c_j(\rho_j/\delta)$.¹² In this case, the estimator is plugged in the covariance matrix as an affine transformation that depends on a normalization that allows to move to the reduced form and fully identifies the first column, so that $R_{1,1} = \delta c_1$ and $R_{1,j} = \beta_j c_j$. More simply, in the first case the estimator reduces to $\beta_j^{IV} = (\rho_j/\delta)$; in the second case the normalization provided by the vector c_j returns the modified estimator $\beta_j^{IV} = \frac{1}{\sqrt{\sigma_{\widetilde{w}_{it}}}}(\rho_j/\delta)$. I will focus on the first case as it provides an interpretation that is akin to the one of a standard IV estimator.

Finally, the following normalizing assumptions will be considered to hold across the rest of the paper.

Assumption 5. (Normalizing assumptions):

- (1) \widetilde{x}_{it} and Z_{it} are stationary,
- (2) $\widetilde{x}_{it} \sim \mathcal{N}(\mathbf{0}_{J+1}, \Sigma)$ and $Z_{it} \sim \mathcal{N}(\mu_Z, \sigma_Z)$.

Notice that assumption 5 is required for several different reasons. Part (1) allows to consider the estimator of first stage and second stage(s) without violations of the Wold theorem and mean that R^{-1} is invertible and part (2) allows the derivative interpretation of δ and ρ_j .

Remark 3. (Why normality?) While normality is not a necessary condition, it has some important properties that are convenient when discussing the estimators. In fact,

et al. (2021) for a full description of the procedure.

¹²Following Gertler and Karadi (2015), consider the partition of the covariance matrix of the residuals $R = [R_1 R_2] = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$, $Q = \frac{r_{21}}{r_{11}} \Sigma_{11} \frac{r'_{21}}{r_{11}} - (\Sigma_{21} \frac{r'_{21}}{r_{11}} + \frac{r_{21}}{r_{11}} \Sigma'_{21}) + \Sigma_{22}$, and $r_{12} r'_{12} = (\Sigma_{21} - \frac{r_{21}}{r_{11}} \Sigma_{11})' Q^{-1} (\Sigma_{21} - \frac{r_{21}}{r_{11}} \Sigma_{11})$. Then, it follows that, to obtain the structural form, $c_j = \sqrt{r_{12}}$ and $R_{1,1} = \delta \cdot c_1$ and $R_{1,k} = \rho_k c_k$.

alternative estimators that make different assumptions about the distribution of Z_{it} and \tilde{x}_{it} may be considered. For example, in the case in which Z_{it} and \tilde{W}_{it} are treatment dummy indicators, the theory goes back to the traditional case of Angrist et al. (1996), and in the case in which Z_{it} and \tilde{W}_{it} are multi-valued, the theory goes back to the cases analyzed by Vytlacil (2002); Heckman and Vytlacil (2001).

Under the assumptions laid out in section 2 it can be shown that

Theorem 1. (PSVAR-IV estimates a ratio of derivatives). *Under assumptions 1;2;3; 5; β^{IV} estimates*

$$\beta_j^{IV} = \frac{\delta \mathbb{E}[\tilde{Y}_j(z^\circ)]/\delta z^\circ}{\delta \mathbb{E}[\tilde{W}(z^\circ)]/\delta z^\circ}.$$

Then, according to theorem 1, β_j^{IV} simply captures the ratio of the effect of moving along different values of z° on \tilde{Y}_j and on \tilde{W} . Therefore, a useful property of impulse response function can be established according to the following theorem.

Theorem 2. (Interpretation of the impulse response functions). *The immediate impulse response function of a shock in \tilde{W} captures*

$$\hat{IRF}_j = \mu - LATE_j = \mathbb{E}[\tilde{Y}_j(\tilde{w}) - \tilde{Y}_j(\tilde{w}') | \tilde{W}(z) = \tilde{w}]$$

representing the difference between the shock being equal to \tilde{w} and \tilde{w}' .

Notice that theorem 2 implies that, considering two different impulse response functions, such as the difference between a 1% and a 0% shock, results in the $\mu - LATE$ that captures the difference of GDP growth for those units that complies with the national spending growth. Hence, the impulse response captures $\mathbb{E}[\tilde{Y}_j(1\%) | \tilde{W}(z) = 1\%]$, the impact of a one percent deviation in regional military spending growth on GDP growth for those units that observed a spending increase.

4 Inference

Instrumental variables are useful only as far as they satisfy the relevance condition (assumption 4). In fact, it is easy to see that, being the IV estimator

$$\beta_j^{IV} = (\rho_j/\delta),$$

weak identification could be tested by the means of a null hypothesis $H_0 : \delta = 0$. Hence, for $\delta \rightarrow 0$, it must be that either $\beta_j^{IV} \rightarrow \infty$ or $\beta_j^{IV} \rightarrow -\infty$ depending on the sign of ρ_j . For some time the general consensus has been to carry two different inferential procedures: one in the first stage, by the means of the Cragg-Donald statistic, frequently defined as first-stage F-statsistic ([Staiger and Stock \(1997\)](#)); and one, separately, in the second stage, by the means of standard confidence intervals or bootstrap. Such procedure heavily relied on the idea that standard confidence intervals possess asymptotically good coverage properties under the alternative hypothesis ($H_1 : \delta \neq 0$).

However, such approach does not cover situations in which the instrument is weak but satisfies independence. Indeed, an approach that generates confidence intervals on β_j^{IV} on the basis of the strength of the instrument, even when the first stage coefficient is near zero, may be preferred to one that may end up discarding interesting research hypothesis on the basis of a weak - but independent - instrument. On the basis of such wisdom, the two step approach may be sub optimal compared to the Anderson-Rubin statistic approach ([Anderson and Rubin \(1949\)](#); [Olea et al. \(2021\)](#); [Andrews et al. \(2019\)](#); [Stock and Yogo \(2002\)](#); [Mikusheva and Poi \(2006\)](#); [Mikusheva \(2010\)](#)). [Mikusheva \(2010\)](#) introduces a procedure for generating confidence sets for the second stage that have good coverage properties even in the null hypothesis case.¹³ [Olea et al. \(2021\)](#) extend the result by introducing a confidence set for the Impulse Response function generated by a SVAR-IV by using the AR statistic.

While it is known that the Anderson-Rubin confidence sets are optimal in the case of one instrument, there is yet to form a consensus about which approach may be preferred

¹³Hence, for $\delta = 0$, the confidence set would be distributed between minus and plus infinity.

in the case of two or more instruments (some recent advancements include the CLR test of Andrews et al. (2006)).

Appendix 7.3 extends the conventional wisdom present in the SVAR-IV field to the PSVAR-IV case using rotations of the Anderson-Rubin statistic and demonstrates the good coverage properties of the AR statistic.

5 Estimation of the dynamic fiscal multiplier

The fiscal multiplier is generally defined as the coefficient a regression with gdp as the dependent variable and government spending as the independent variable. Such measure is of great interest because of its policy relevance: a relatively large fiscal multiplier is often times evoked by governments as the reason to increase spending¹⁴.

The aggregate fiscal multiplier is generally computed using vector autoregressions (Romer and Romer (2010); Blanchard and Perotti (2002)) or local projections (Ramey and Zubairy (2018)). Normally, the aggregate fiscal multiplier was found to be rarely above one.

A local fiscal multiplier could be preferred to the aggregate fiscal multiplier for several reasons. First, the assumptions required to obtain an unbiased estimand are less restrictive than their aggregate counterpart. Indeed, the computation of a national aggregate fiscal multiplier often poses some credibility issues due to the unreliability of the underlying assumptions. The fiscal policy literature has therefore explored the quantification of the impact of a fiscal expansion on GDP by using more granular and localized data, either at the state or regional-level. Second, the open economy multiplier can be potentially more interesting to central bankers because, by using state heterogeneity, it is essentially independent of monetary policy, as overnight rates are fixed for all states.

The recent emergence of this literature has generated different relevant contributions that seem to indicate a regional fiscal multiplier of about 1.5 (Farhi and Werning (2016); Nakamura and Steinsson (2014); Shoag et al. (2010); Chodorow-Reich (2019)). To the

¹⁴For example, Nakamura and Steinsson (2014) observes that the American Recovery and Reinvestment Act (ARRA) was justified on the basis of large estimates of the fiscal multiplier.

best of my knowledge, there is currently little work done in estimating the dynamic *regional* fiscal multiplier with the same type of external instrument approach that has characterized the aggregate data literature. Perhaps closer to the main idea of this paper is [Dupor et al. \(2023\)](#), which estimates the dynamic regional fiscal multiplier using a model to frame the impact of the ARRA. Yet, the ARRA is representative of a particular context of the US economy of low inflation and low interest rates, and may not be representative of different state dependencies.

Motivated by the lack of empirical evidence at the intersection of the two streams of literature, I estimate a regional fiscal multiplier for the US using aggregate national military spending as an instrument for the innovations of regional military spending. Two cautions are however invited to the reader. First, the fiscal multiplier identified using military spending data is particularly useful, but may not be representative of a generic spending multiplier. It is useful as it is inherently a measure of direct spending of the US government ([Nakamura and Steinsson \(2014\)](#)); but it is not representative as it does not include *all* the government spending ([Koo et al. \(2023\)](#)).

In the case of panel vector autoregressions, the dynamic regional fiscal multiplier can be estimated by simply by running a PVAR estimation on the vector

$$x_{it} = \left(\frac{\exp_{it} - \exp_{it-1}}{\text{gdp}_{it-1}}, \frac{\text{gdp}_{it} - \text{gdp}_{it-1}}{\text{gdp}_{it-1}} \right)'$$

Here, the issue of endogeneity arises because the contemporaneous innovations of fiscal expanses growth may not be thought as exogenous with respect to the contemporaneous innovations of GDP growth.

The leading assumption for the case of PVAR therefore is that the United States do not embark on military buildups because states that receive a disproportionate amount of military spending are doing *more poorly than before* relative to other states. To exploit this assumption, I use data from the US extracted from the electronic database of DD-350 military procurement forms available from the US Department of Defense by [Nakamura and Steinsson \(2014\)](#), which includes military spending for equipment of 10000\$ or more

	MBIC	MAIC	MQIC
$p = 1$	-24.08	7.00	-5.35
$p = 2$	-23.02	-7.48	-13.66

Table 1: MAIC, BBIC, MHIQ tests.

in the period 1966-1984 and above 100000\$ in the period until 2006.¹⁵ The rest of the analysis follows Nakamura and Steinsson (2014) fairly closely: the data is at a yearly frequency, and region refers to the aggregation of different states that are close and not densely populated, resulting in 10 different macro-regions and 39 different years. Differently from the original paper, however, the main estimation makes use of the variable's growth with respect to the previous year, rather than the previous two years; and the Bartik/shift-share instrument is given a preference over using 10 different instruments (one for each state). The reason I made such choice is that conventionally time series regressions are framed in terms of growth with respect to the previous period, and a one dimensional instrument is known to have an optimal confidence set, whereas the case of multiple instruments may provide less reliable confidence sets.¹⁶

The model is estimated as follows. First, I use a one-lag model as suggested by the MAIC, MBIC, MHIQ from table 1. The residuals plotted in Figure 1. From the figure, there appears to be no indication of residual autocorrelation. This is confirmed by the regression coefficients obtained by regressing the residuals against their lags. The coefficients, being not statistically different from zero, do not seem to suggest to reject the null hypothesis of a statistically significant relationship between the residuals and their lags. Finally, I turn to the assumption of normality of the model, discussed in assumption 5. In fact, violations of assumption 5(iii) would suggest that non-parametric estimators could be preferred over the ones implied by the 2sls utilized for the IV regression because of the assumption that are required from parametric estimators. The histograms displaying

¹⁵Unfortunately, the data is not updated any further.

¹⁶In the appendix, I show that alternative formulations using growth with respect to two previous periods may change the results slightly. While the quantities tend to be similar in the impulse response, the mechanism by which fiscal expansions tend to be associated with an increase in output in the following year is by the means of a large output autocorrelation, rather than a direct correlation between output and past expanses.

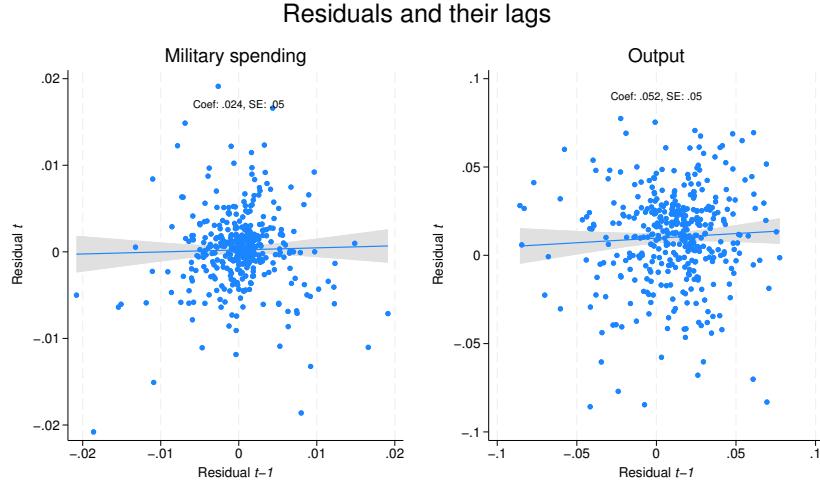


Figure 1: Regression of the residuals and their lags.

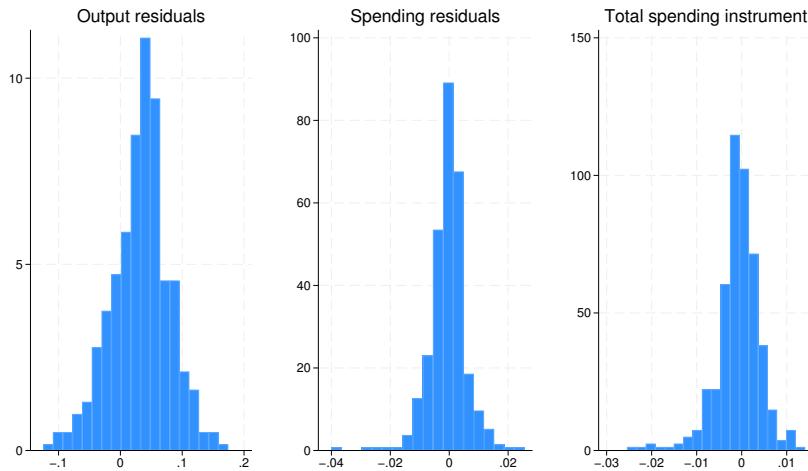


Figure 2: Distribution of the residuals and of the national military spending instruments.

the error's distribution in figure 2 seem to indicate that the residuals may be normally distributed and therefore continuously differentiable with appropriate weights, indicating the adequateness of the normality assumption.¹⁷

The results from the IV regression are instead displayed in Table 2. The coefficient from the first stage is statistically significantly different from zero, and the first stage F-statistic is above the commonly advocated threshold level of 10 (see [Staiger and Stock \(1997\)](#)). Moreover, the AR statistic is above the $\chi_{1,1-\alpha}$ critical value for $\alpha = .05$, sug-

¹⁷Several statistical tests (such as [Shapiro and Wilk \(1965\)](#); [Shapiro and Francia \(1972\)](#)) with the null hypothesis of normality fail to reject the null, indicating that the residuals may indeed be normally distributed.

FIRST STAGE		SECOND STAGE	
	<i>State spending</i>		<i>GDP growth</i>
<i>Total spending</i>	0.46	<i>State spending</i>	1.74
<i>Standard CI</i>	[0.37 0.54]	<i>Anderson-Rubin CS</i>	[0.53 3.05]
<i>Fist stage F statistic</i>	120.86	<i>Anderson-Rubin statistic</i>	92.63

Table 2: The first column of this table reports the result of the regression of state spending innovations on total spending, with standard confidence intervals and the first stage F-statistic. The second column of the table instead reports the results of the 2sls of GDP growth innovations on the instrumented state spending innovation, with the confidence sets built using the Anderson-Rubin statistic and the Anderson-Rubin statistic itself. In both cases, the confidence interval and the confidence set are at the 95% level.

	$\frac{\exp_{it-1} - \exp_{it-2}}{gdp_{it-2}}$	$\frac{gdp_{it-1} - gdp_{it-2}}{gdp_{it-2}}$
$\frac{\exp_{it} - \exp_{it-1}}{gdp_{it-1}}$	-0.10 [-0.28, 0.06]	-0.03*** [-0.05, -0.01]
$\frac{gdp_{it} - gdp_{it-1}}{gdp_{it-1}}$	0.80*** [0.20, 1.41]	0.34*** [0.21, 0.46]

Table 3: Autoregressive coefficients of the PVAR. The confidence interval represent are set at the 95% level.

gesting that the results may be statistically significantly different from zero.

Finally, consider figure 3, which displays the impulse response functions of a 1% shock in fiscal spending growth. The results are similar to the ones of the literature, suggesting a value of the fiscal multiplier of approximately ~ 1.7 in the first period¹⁸. However, the dynamic fiscal multiplier displays an interesting feature, as it appears that the impact of a change in the fiscal spending in year t results in a corresponding increase in output growth by ~ 1.5 . To better highlight the mechanism by which such response happens, table 3 displays the AR coefficients. The high correlation between GDP growth and fiscal policy in the previous period is the main mechanism by which the fiscal multiplier can result in a GDP growth that may last for more than one year. Moreover, fiscal policy tends to not be particularly correlated with past fiscal policy or output, resulting in a response close to zero in the second horizon.

¹⁸Notice that, by definition, the IRF on impact is the second stage regression in table 2.

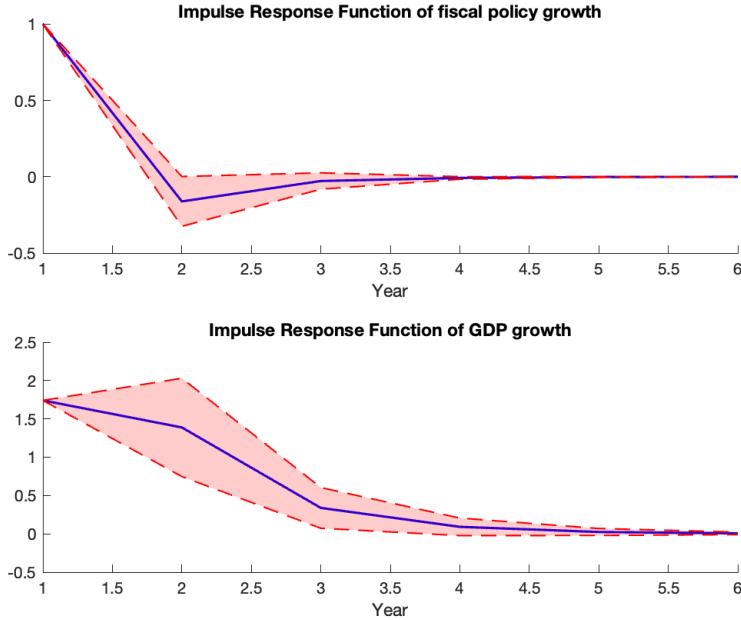


Figure 3: Impulse response functions of regional military spending and regional GDP growth to a 1% shock in regional military spending. The confidence sets are built using the Anderson-Rubin test statistic developed in section 4 at the 95% level.

6 Conclusions

This paper discussed the causal interpretation of panel vector autoregressions identified by the means of external instruments. The IRF generated by a PVAR can estimate a LATE representing the difference between the outcome variable under a treatment and no treatment status for the compilers. However, such LATE needs to be read differently from the panel linear regression literature, as it refers to the residuals and emerges as a counterfactual assignment of different predictions, such as a 1% shock versus a 0% shock. I have discussed under which assumptions the LATE may be captured: independence, exclusion, and monotonicity. Some drawbacks of the proposed identification scheme include the severity of the parametric linear nature of the monotonicity assumption.

Moreover, I discussed the best approaches to conduct inference in a PVAR identified using external instrument. In appendix 7.3 I showcase the good small sample properties of the AR confidence sets calibrating a simulation on the basis of the dataset from the application.

Finally, I have applied these tools to the estimation of a dynamic regional fiscal

multiplier for the United States, a quantity that has been rarely targeted by the literature. My empirical findings suggest that the dynamic regional fiscal multiplier may be above one in the second period, indicating some possibly longer term effects of fiscal expansions on GDP growth.

Future researchers are invited to develop two points. First, the μ – LATE interpretation of the PSVAR-IV relies on an underlying linearity assumption. Yet, non-parametric estimators, which could potentially alleviate the linearity assumption, are never utilized in the SVAR nor the PSVAR literature. If the data utilised is sufficiently large, such methods could be further explored. Second, because the IV literature is still uncertain about which statistics to use when dealing with multiple instruments, the inference issue of overidentification naturally carry to SVAR-IV and PSVAR-IV. Hence, future researchers should properly discuss the unreliability of confidence sets in such cases and possibly implement novel methodologies with better coverage properties.

References

- Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46 – 63.
- Andrews, D. W., Moreira, M. J., and Stock, J. H. (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74(3):715–752.
- Andrews, I., Stock, J., and Sun, L. (2019). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*, 11:727–753.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.
- Antonelli, J., Mealli, F., Beck, B., and Mattei, A. (2023). Principal stratification with continuous treatments and continuous post-treatment variables. *arXiv preprint arXiv:2309.14486*.
- Bartik, T. J. (1991). Who benefits from state and local economic development policies?
- Blanchard, O. and Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*, 117(4):1329–1368.
- Bojinov, I. and Shephard, N. (2019). Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*.
- Brignone, D., Franconi, A., and Mazzali, M. (2023). Robust impulse responses using external instruments: the role of information. *arXiv preprint arXiv:2307.06145*.
- Bruns, M. and Keweloh, S. A. (2024). Testing for strong exogeneity in proxy-vars. *Journal of Econometrics*, 245(1-2):105876.

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

Dupor, B., Karabarbounis, M., Kudlyak, M., and Saif Mehkari, M. (2023). Regional consumption responses and the aggregate fiscal multiplier. *Review of Economic Studies*, 90(6):2982–3021.

Farhi, E. and Werning, I. (2016). Fiscal multipliers: Liquidity traps and currency unions. In *Handbook of macroeconomics*, volume 2, pages 2417–2492. Elsevier.

Fieller (1944). A fundamental formula in the statistics of biological assay, and some applications. *Quarterly Journal of Pharmacy and Pharmacology*, 17:117–123.

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

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Granger, C. W. J. and Newbold, P. (2014). *Forecasting economic time series*. Academic press.

Heckman, J. J. and Vytlacil, E. (2001). Policy-relevant treatment effects. *American Economic Review*, 91(2):107–111.

Kolesár, M. and Plagborg-Møller, M. (2024). Dynamic causal effects in a nonlinear world: the good, the bad, and the ugly. *arXiv preprint arXiv:2411.10415*.

Koo, B., Lee, S., and Seo, M. H. (2023). What impulse response do instrumental variables identify?

Menchetti, F. and Bojinov, I. (2022). Estimating the effectiveness of permanent price reductions for competing products using multivariate Bayesian structural time series models. *The Annals of Applied Statistics*, 16(1):414 – 435.

- Menchetti, F., Cipollini, F., and Mealli, F. (2022). Combining counterfactual outcomes and ARIMA models for policy evaluation. *The Econometrics Journal*, 26(1):1–24.
- Mertens, K. and Ravn, M. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, 103(4):1212–1247.
- Mertens, K. and Ravn, M. (2014). A reconciliation of svar and narrative estimates of tax multipliers. *Journal of Monetary Economics*, 68:S1–S19.
- Mikusheva, A. (2010). Robust confidence sets in the presence of weak instruments. *Journal of Econometrics*, 157(2):236–247.
- Mikusheva, A. and Poi, B. P. (2006). Tests and confidence sets with correct size when instruments are potentially weak. *The Stata Journal*, 6(3):335–347.
- Nakamura, E. and Steinsson, J. (2014). Fiscal stimulus in a monetary union: Evidence from us regions. *American Economic Review*, 104(3):753–792.
- Nakamura, E. and Steinsson, J. (2018). Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86.
- Olea, J. L. M., Stock, J. H., and Watson, M. W. (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, 225(1):74–87.
- Pala, R. (2025). The causal interpretation of panel vector autoregressions. *arXiv preprint arXiv:2510.23540*.
- Rambachan, A. and Shephard, N. (2021). When do common time series estimands have nonparametric causal meaning? *Working Paper*.
- Ramey, V. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from us historical data. *Journal of Political Economy*, 126(2):850 – 901.

- Romer, C. D. and Romer, D. H. (2010). The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American economic review*, 100(3):763–801.
- Shapiro, S. S. and Francia, R. (1972). An approximate analysis of variance test for normality. *Journal of the American statistical Association*, 67(337):215–216.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.
- Shoag, D. et al. (2010). The impact of government spending shocks: Evidence on the multiplier from state pension plan returns. *unpublished paper, Harvard University*.
- Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.
- Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948.
- Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression.
- Vytlačil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341.
- Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4):478–486.

	MBIC	MAIC	MQIC
$p = 1$	-24.62	36.16	11.91
$p = 2$	-33.22	12.37	-5.82
$p = 3$	-29.53	0.87	-11.26
$p = 4$	-16.81	-1.61	-7.68

Table 4: MAIC, BBIC, MHIQ tests.

FIRST STAGE		SECOND STAGE	
	<i>State spending</i>		<i>GDP growth</i>
<i>Total spending</i>	0.49	<i>State spending</i>	2.27
<i>Standard CI</i>	[0.39 0.60]	<i>Anderson-Rubin CS</i>	[0.73 3.97]
<i>Fist stage F statistic</i>	89.50	<i>Anderson-Rubin statistic</i>	8.13

Table 5: The first column of this table reports the result of the regression of state spending innovations on total spending, with standard confidence intervals and the first stage F-statistic. The second column of the table instead reports the results of the 2sls of GDP growth innovations on the instrumented state spending, with the confidence sets built using the Anderson-Rubin statistic and the Anderson-Rubin statistic itself. In both cases, the confidence interval and the confidence set are at the 95% level. In this case, the growth rates are computed using the growth compared to 2 years before, and the PVAR is estimated using 2 lags.

7 Appendix

7.1 Appendix A: Other empirical results

I propose a robustness check based on alternative growth rate measures. This is because Nakamura and Steinsson (2014) prefer two years growth rate to growth rates with respect to the previous year. Using growth rates with respect to two years before I obtain the MBIC-MAIC-MHQ in table 4. Different tests appear to suggest different lag selections, but I will make use of a 2 lag specification that minimizes the MBIC.

In this case, the results from the 2sls are reported in table 5. The table suggests a slightly larger fiscal multiplier on impact compared to the one found in the main specification. Finally, the impulse response is displayed in figure 4. From the figure, it appears as though the impact of a shock in fiscal policy on output may be larger in the second period compared to the main specification provided in the paper. This feature appears to be almost entirely driven by a relatively large autocorrelation coefficient of order 1 of

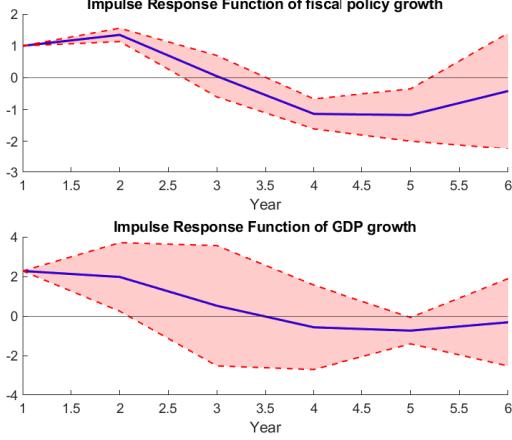


Figure 4: Impulse response functions of regional military spending and regional GDP growth to a 1% shock in regional military spending. The confidence sets are built using the Anderson-Rubin test statistic developed in section 4 at the 95% level.

$t - 1$		$t - 2$		
$\frac{\exp_{it-1} - \exp_{it-3}}{gdp_{it-3}}$	$\frac{gdp_{it-1} - gdp_{it-3}}{gdp_{it-3}}$	$\frac{\exp_{it-2} - \exp_{it-4}}{gdp_{it-4}}$	$\frac{gdp_{it-2} - gdp_{it-4}}{gdp_{it-4}}$	
$\frac{\exp_{it} - \exp_{it-2}}{gdp_{it-2}}$	0.63*** [0.47, 0.77]	-0.03** [-0.04, -0.012]	-0.21*** [-0.35, -0.80]	-0.003 [-0.01, 0.01]
$\frac{gdp_{it} - gdp_{it-2}}{gdp_{it-2}}$	0.32 [-0.32, 0.96]	0.88*** [0.77, 0.98]	-0.53 [-1.17, 0.11]	-0.52*** [-0.42, -0.62]

Table 6: Autoregressive coefficients of the PVAR estimated using growth levels with respect to two years before and two lags. The confidence interval are set at the 95% level.

local military spending. In fact, while in the main body of the paper such AR is negative and close to zero, in this model it is estimated to be about .6, as displayed in table 6. Such movement is however counteracted by a contraction predicted by a negative coefficient in the second lag. Overall, the interpretation of the impulse response function is slightly different from the one in the main specification because the channel by which gdp is supposed to increase is also largely driven by its own autocorrelation. In the end, the impact still appears to be positive and statistically significant for the contemporaneous impact and the following period, and then is zero after. At the same time, the impact of a fiscal policy spending shock implies a future decline of fiscal policy. Therefore, on broader terms, this specification confirms a positive and statistically significant impact of fiscal expansion to GDP in the following year.

7.2 Appendix B: Proofs regarding identification

Lemma 1. Consider (without loss of generality) the case of a system of the kind $x_{it} = (W'_{it}, Y'_{1,it}, \dots, Y'_{J,it})'$. Let me define \mathbf{x} as the matrix containing each x_{it} and $\mathbf{x}_{2:J}$ as the partition that includes all the outcome variables, which are ordered from the second to the last position. Moreover, $\boldsymbol{\omega}$ will be defined as the matrix containing all the lags, i.e. $\omega_{it} = (W'_{it-1}, Y'_{j,it-1}, \dots, Y'_{J,it-1})'$. The residuals of the VAR are $\tilde{\mathbf{x}} = (\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}\boldsymbol{\omega}'\mathbf{x}$. Then, it is possible to consider the partition $\tilde{\mathbf{x}}_{2:J}$ as the one containing the residuals of the outcome variables, and $\tilde{\mathbf{x}}_1$ as the one containing the residuals of the first column. Then, the 2sls estimator becomes

$$\beta_j^{IV} = ((Z'Z)^{-1}Z'\tilde{\mathbf{x}}_{2:J})/((Z'Z)^{-1}Z'\tilde{\mathbf{x}}_1).$$

Here, in the case in which Z is continuously distributed, $\beta_j^{IV} = \frac{\delta\mathbb{E}[\tilde{Y}_{j,it}|Z_{it}=z^\circ]/\delta z^\circ}{\delta\mathbb{E}[W_{it}|Z_{it}=z^\circ]/\delta z^\circ}$.

Proof. **Proof of theorem 1.** Recall that, from lemma 1,

$$\beta_j^{IV} = \rho_j/\gamma = \frac{cov(\tilde{Y}_j, Z)}{var(Z)}/\frac{cov(\tilde{W}, Z)}{var(Z)}.$$

Then, the proof follows as in theorem 4.3 of [Pala \(2025\)](#) and is similar to [Yitzhaki \(1996\)](#). Because the estimator is fundamentally composed by ρ_j and γ , it can be decomposed in two components which capture a similar estimands. First, let me consider $cov(\tilde{Y}_j, Z)$:

$$\begin{aligned} cov(\tilde{Y}_j, Z) &= \mathbb{E}[(\tilde{Y}_j - \mathbb{E}(\tilde{Y}_j))(Z - \mathbb{E}(Z))] \\ &= \mathbb{E}[(\tilde{Y}_j(Z - \mathbb{E}(Z)))] \\ &= \mathbb{E}[(\mathbb{E}[\tilde{Y}_j|Z])(Z - \mathbb{E}(Z))] \\ &= \int (z^\circ - \mathbb{E}(Z))g(z^\circ)f_Z(z^\circ)dz^\circ. \end{aligned}$$

Here the first equality holds because of the law of the covariance, the second holds because the innovations are assumed to be zero mean for each of the outcome variables, the third holds by the law of total expectations, and the last holds by rewriting the expected value as an integral and defining $g(z^\circ) = \mathbb{E}[\tilde{Y}_j|Z = z^\circ]$. Defining $\nu'(m) = (z^\circ - \mathbb{E}[\tilde{Y}_j])f_Z(m)$,

$v(m) = \int_{-\infty}^Z (m - \mathbb{E}[Z]) f_Z(m) dm$ and $u(z^\circ) = g(z^\circ)$ I can apply integration by parts to obtain

$$Cov(\tilde{Y}_j, Z) = \int_{-\infty}^Z (m - \mathbb{E}[Z]) f_Z(m) dm g(z^\circ) - \int_{-\infty}^{\infty} \left(\int_{-\infty}^Z (m - \mathbb{E}[Z]) f_Z(m) dm g'(z^\circ) dz^\circ \right). \quad (4)$$

Notice that the first part converges to zero if the variance of Z exists, and changing the sign to the second part we obtain

$$\begin{aligned} Cov(\tilde{Y}_j, Z) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^Z (\mathbb{E}[Z] - m) f_Z(m) dm g'(z^\circ) dz^\circ \right) \\ &= \int_{-\infty}^{\infty} (\mathbb{E}[Z] F_Z(z^\circ) - \theta_Z(z^\circ)) g'(z^\circ) dz^\circ, \end{aligned}$$

where the first equality holds by changing the sign of the second part of 4, the second holds by substituting the definition of $\theta_Z(z^\circ) = \int_{-\infty}^Z m f_Z(m) dm$.

And the denominator is $var(Z) = \sigma_Z^2$. Therefore,

$$\frac{cov(\tilde{Y}_j, Z)}{var(Z)} = \frac{\int_{-\infty}^{\infty} (\mathbb{E}[Z] F_Z(z^\circ) - \theta_Z(z^\circ)) g'(z^\circ) dz^\circ}{\sigma_Z^2},$$

which is equivalent to the one in the theorem by using the definition of the weights $q(z^\circ) = \frac{1}{\sigma_Z^2} \int_{-\infty}^{\infty} (\mathbb{E}[Z] F_Z(z^\circ) - \theta_Z(z^\circ)) dz^\circ$. Now consider the form

$$\frac{cov(\tilde{Y}_j, Z)}{var(Z)} = \int q(z^\circ) g'(z^\circ) dz^\circ.$$

Substituting $q(z^\circ) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z^\circ} me^{-m^2/2} dm = \frac{1}{\sqrt{2\pi}} e^{-m^2/2}$ inside the definition of γ_k , I obtain the following

$$\begin{aligned} \frac{cov(\tilde{Y}_j, Z)}{var(Z)} &= \int \frac{1}{\sqrt{2\pi}} e^{-m^2/2} g'(z^\circ) dz^\circ \\ &= g'(z^\circ) \int \frac{1}{\sqrt{2\pi}} e^{-m^2/2} dz^\circ \\ &= g'(z^\circ), \end{aligned}$$

where the first equality is true by substituting the value of the weights $q(z^\circ)$, the second by the fact that the density of $g'(z^\circ)$ does not depend on z° , and the last by the laws of integration of normal variables which are satisfied by Z according to assumption 5(iii). Then, $\frac{\text{cov}(\tilde{Y}_j, Z)}{\text{var}(Z)}$ can be expressed as:

$$\frac{\delta \mathbb{E}[\tilde{Y}_j(z^\circ) | Z = z^\circ]}{\delta z^\circ} = \frac{\delta \mathbb{E}[\tilde{Y}_j(z^\circ)]}{\delta z^\circ} + \frac{\text{cov}(\tilde{Y}_j(z^\circ), 1\{Z = z^\circ\})}{\text{var}(1\{Z = z^\circ\})}.$$

By assumption 1, $\text{cov}(\tilde{Y}_j(z^\circ), 1\{Z = z^\circ\}) = 0$, which returns:

$$\frac{\delta \mathbb{E}[\tilde{Y}_j(z^\circ) | Z = z^\circ]}{\delta z^\circ} = \frac{\delta \mathbb{E}[\tilde{Y}_j(z^\circ)]}{\delta z^\circ}$$

Then, applying the same steps to $\frac{\text{cov}(\tilde{W}, Z)}{\text{var}(Z)}$, we obtain the result in the theorem. \square

Proof. **Proof of theorem 2.** The immediate period impulse response function is computed as $(1, 0'_J)B$. Therefore, each contemporaneous IRF is simply defined as $\text{IRF}_j = c \cdot \beta_j^{IV}$. Starting from theorem 1, it is therefore easily possible to recast the IRF

$$\frac{\delta \mathbb{E}[\tilde{Y}_j(z^\circ)]}{\delta z^\circ} / \frac{\delta \mathbb{E}[\tilde{W}_j(z^\circ)]}{\delta z^\circ}$$

as the difference between a shock z and a shock z'

$$\mathbb{E}[\tilde{Y}_j(z) - \tilde{Y}_j(z')]/\mathbb{E}[\tilde{W}(z) - \tilde{W}(z')]$$

where $\tilde{Y}_j(z)$ indicates the IRF under the assignment of a shock and $\tilde{Y}_j(z')$ under the assignment of a different shock. Essentially, as long as the linearity assumptions of ρ_j and γ are satisfied, the counterfactual assignment imposed when constructing the impulse response function can be interpreted as follows:

$$\mathbb{E}[\tilde{W}(z) - \tilde{W}(z')] = \mathbb{E}[\tilde{W}(z)] = \mathbb{P}[\tilde{W}(z) = \tilde{w}].$$

Moreover,

$$\begin{aligned}\mathbb{E}[\tilde{Y}_j(z)] &= \mathbb{E}[\tilde{Y}_j(\tilde{W}(z), z)] = \mathbb{E}[\tilde{Y}_j(\tilde{w}, z) | \tilde{W}(z) = \tilde{w}] \mathbb{P}[\tilde{W}(z) = \tilde{w}] \\ &\quad + \mathbb{E}[\tilde{Y}_j(\tilde{w}', z) | \tilde{W}(z) = \tilde{w}'](1 - \mathbb{P}[\tilde{W}(z) = \tilde{w}])\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\tilde{Y}_j(z')] &= \mathbb{E}[\tilde{Y}_j(\tilde{w}', z')] = \mathbb{E}[\tilde{Y}_j(\tilde{w}', z') | \tilde{W}(z) = \tilde{w}] \mathbb{P}[\tilde{W}(z) = \tilde{w}] \\ &\quad + \mathbb{E}[\tilde{Y}_j(\tilde{w}', z') | \tilde{W}(z) = \tilde{w}'](1 - \mathbb{P}[\tilde{W}(z) = \tilde{w}])\end{aligned}$$

Therefore, the numerator becomes equivalent to the classic form of an ITT as

$$\begin{aligned}\mathbb{E}[\tilde{Y}_j(z)] - \mathbb{E}[\tilde{Y}_j(z')] &= \mathbb{E}[\tilde{Y}_j(\tilde{w}, z) - \tilde{Y}_j(\tilde{w}', z') | \tilde{W}(z) = \tilde{w}] \mathbb{P}[\tilde{W}(z) = \tilde{w}] + \\ &\quad \mathbb{E}[\tilde{Y}_j(\tilde{w}, z) - \tilde{Y}_j(\tilde{w}', z') | \tilde{W}(z) = \tilde{w}'](1 - \mathbb{P}[\tilde{W}(z) = \tilde{w}])\end{aligned}$$

and the estimator becomes

$$\frac{\mathbb{E}[\tilde{Y}_j(\tilde{w}, z) - \tilde{Y}_j(\tilde{w}', z') | \tilde{W}(z) = \tilde{w}] \mathbb{P}[\tilde{W}(z) = \tilde{w}]}{\mathbb{P}[\tilde{W}(z) = \tilde{w}]} = \mathbb{E}[\tilde{Y}_j(\tilde{w}) - \tilde{Y}_j(\tilde{w}') | \tilde{W}(z) = \tilde{w}]$$

which concludes the proof.

□

7.3 Appendix C: Proofs and simulations regarding inference

In this appendix, I show that the case of PSVAR-IV is not too dissimilar from the case of SVAR-IV. In this case, the assumptions required for the nominal good coverage of the AR statistic are simply related to the instrument's exogeneity and the convergence of the covariance matrix of the residuals as well as the covariance between $\tilde{W}_{i,t}$ and $Z_{i,t}$.

Proposition 1. Let $\text{CS}^{\text{AR}}(1 - \alpha)$ be the AR set. \mathcal{P}_T is the probability distribution of $\{x_{i,t}, Z_{i,t}\}_{i=1, t=1}^{I, T}$. δ is the covariance of $(Z_{i,t}, \tilde{W}_{i,t})$. Then, suppose:

- (i) Assumption 1,..,3 are satisfied, which implies $\mathbb{E}[\tilde{Y}_{j,i,t}, Z_{i,t}] = 0$ for all $j = 1, \dots, J$
- (ii) $\delta_T \rightarrow \delta$
- (iii) $\Sigma_T \rightarrow \Sigma$

Then, $\lim_{T \rightarrow \infty} \mathcal{P}_T(\lambda_{k,s} \in \text{CS}^{\text{AR}}(1 - \alpha)) = 1 - \alpha$.

Proof. **Proof of proposition 1.**¹⁹ Let $\lambda_{k,s}$ denote the true impulse response coefficient of variable s to a shock in the variable k and consider the statistic $G_T = [\sqrt{T}(e'_s C_k(\hat{\Phi}_T) - \lambda_{k,s} e'_s)]$ where e'_s is a vector that slices an identity matrix I_n , so that it selects the impulse response function of variable s . Moreover, $C_k(\hat{\Phi}_T)$ represents the moving average representation of the autoregressive coefficient $\hat{\Phi}_T$ and $\lambda_{k,i}$ represents the impulse response function $\lambda_{k,s} = e'_s C_k(\Phi) \Gamma / e'_s \Gamma$, with $\Gamma = (\delta', e'_s)'$ for all $s = 1, \dots$ and stacks the projections of the instrumented and outcome variables on $Z_{i,t}$. Then, the AR statistic is by definition

$$\mathcal{P}_T(\lambda_{k,s} \in \text{CS}_T^{\text{AR}}(1 - \alpha)) = \mathcal{P}_T(G_T \leq z_{1-\alpha, 2}^2 \hat{\sigma}_T^2(\lambda_{k,s}))$$

where $\hat{\sigma}_T^2(\lambda_{k,s})$ is the estimator of the asymptotic variance of G_T . Then, the covariance matrix of the residuals Σ is positive definite by assumption 5 and therefore $\sigma^2(\lambda_{k,s}) \neq 0$. Then,

$$G_T^2 / \hat{\sigma}_T^2(\lambda_{k,s}) \xrightarrow{d} \chi_1^2$$

¹⁹The proof follows similarly to Olea et al. (2021) with the difference that it refers to panel data rather than a time series. It is therefore hereby reported to highlight their differences.

follows from assumption 1, 2, 3. Then, $\lim_{t \rightarrow \infty} \mathcal{P}_T(\lambda_{k,s} \in \text{CS}_T^{\text{AR}}(1 - \alpha)) = 1 - \alpha$. \square

Remark 4. The asymptotic properties of the CS_T^{AR} only require $t \rightarrow \infty$, but not $i \rightarrow \infty$. This means that the convergence can happen for a fixed number of units, as long as t goes to infinity.

Remark 5. Notice that assumption 4 is not required for the convergence of the CS_T^{AR} statistic. Hence, the validity of the confidence set holds even if, say, δ is small and close to zero. In such case, the F-test may be small but the CS_T^{AR} will be valid.

Proposition 2. Let $\text{CS}^{\text{AR}}(1 - \alpha)$ be the AR set and $\text{CS}^{\text{Plug-in}}(1 - \alpha)$ be the plug-in estimator of Olea et al. (2021). Here d_H is the probability distribution of $H = \begin{bmatrix} e_s' C_k(\Phi) \Gamma_T \\ e_1' \Gamma_T \end{bmatrix}$, so that \hat{H}_T is the plug-in estimator of H_T constructed by replacing (Φ, Γ_T) with $(\hat{\Phi}, \hat{\Gamma}_T)$, so that $\sqrt{T}(\hat{H}_T - H_T) \rightarrow \eta \sim N(0, \Sigma)$

- (i) Assumption 1,..,3 are satisfied, which implies $\mathbb{E}[\tilde{Y}_{j,i,t}, Z_t] = 0$ for all $j = 1, \dots, J$
- (ii) $\delta_T \rightarrow \delta$
- (iii) $\Sigma_T \rightarrow \Sigma$
- (iv) $\hat{\sigma}_{T,k,s}^2 \xrightarrow{p} \sigma_{T,k,s}^2$

Then, $\sqrt{T}d_H(\text{CS}^{\text{AR}}(1 - \alpha), \text{CS}^{\text{Plug-in}}(1 - \alpha)) \xrightarrow{p} 0$

Proof. **Proof of proposition 2.** The proof follows from Olea et al. (2021), Appendix A2.2. The convergence rate here is also \sqrt{T} by simply relying on the convergence properties of PVARs under a fixed N asymptotics. \square

Finally, the rate of convergence of the coefficient may depend on several factors, including unit-heterogeneity. To analyse whether unit heterogeneity may impact the coverage properties of the AR set, I set up a Monte Carlo simulation that is parametrized according to the observable data. In this case, I consider $T = 39$ and $N = 10$ as the original dataset. The data is generated according to the estimated Φ and Σ . Finally, the matrix of the impulse response coefficients R is set up to be $b/\sqrt{b'\Sigma b}$ in the first column, where $b = (11)'$. The remaining columns of are chosen to satisfy $RR' = \Sigma$. The external

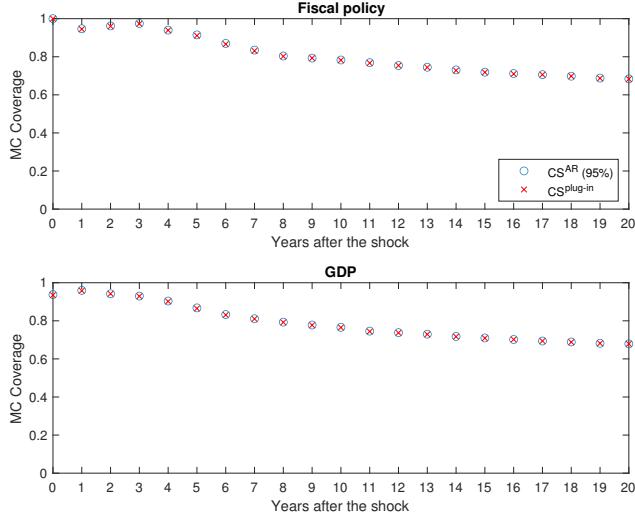


Figure 5: Monte Carlo simulations of the coverage of the IRF.

instrument is set to be

$$Z_t = \mu_Z + \gamma \widetilde{W}_{i,t} + \sigma_Z \nu_t$$

where μ_Z is estimated according to the mean of the instrument (approximately 0) and σ_Z is the variance of the aggregate fiscal policy instrument (0.005). The concentration parameter $((TN)\alpha)^2/Cov(\widetilde{W}_{i,t}, Z_t)$ is computed to be about 204. Hence, figure 5 indicates a good coverage (above 95) from the impulse response on impact and on the following periods. In this case, because of the informativeness of the panel component, the convergence appears to be particularly fast even in a small T scenario (39). However, the coverage tends to decline as the impulse response function goes to more and more horizons. This features tends to happen mainly due to the small point precision of an impulse response function computed using an AR(1) process. This is not a property shared by the cumulative impulse response, which instead tends to have good coverage properties even for large horizons of the IRF. This feature can be seen in figure 6.

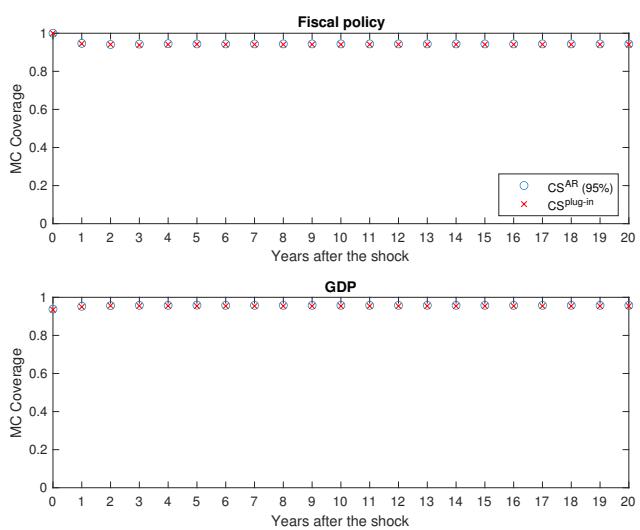


Figure 6: Monte Carlo simulations of the coverage of the cumulative IRF.