Identification and Estimation of Heterogeneous Dynamic Causal Effects using Local Projection Instrumental Variable

Bonsoo Koo*
Monash University

Seojeong Lee†
University of New South Wales

Myunghwan Seo‡
Seoul National University

December 1, 2020

1 Introduction

Instrumental variables (IV) provide a convenient but powerful tool to overcome the endogeneity problem in identification and estimation of the parameter of interest in economic models. In the microeconometrics literature, using an exogenous variable correlated with the endogenous variable but is uncorrelated with the unobserved error term in the main equation as an IV is common practice. Although IV regression is less common in the macroeconometrics literature, there has been increasing attention to using an external shock as an IV. The local projections (LP) is one of such approaches. The LP method was proposed by Jordà (2005) to compute the impulse-response functions (IRF) without fully specifying the law of motion of the underlying multivariate system, as an alternative to the commonly used vector autoregressions (VAR). It has been extended to models with endogeneity using IV in Jordá and Taylor (2015), Ramey and Zubairy (2018), and Stock and Watson (2018). Following the literature, we call this IV regression of the LP models LP-IV.

A standard setting in the modern microeconometrics literature on IV is to allow for heterogeneity in the treatment effects. In a series of seminal papers, Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist, Imbens, and Rubin (1996) investigate identification and estimation of causal effects using IV under treatment effect heterogeneity. An important finding is that the IV estimand identifies the average treatment effect (ATE) of those who receive treatment because of the IV, which they call the local average treatment effect (LATE). Heckman and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2006) show LATE can be written as a weighted average of the

*bonsoo.koo@monash.edu
†jay.lee@unsw.edu.au
‡myunghseo@snu.ac.kr

marginal treatment effects and provide a more general framework to characterize policy-relevant treatment effects.

Unfortunately, these important findings have not been fully incorporated into the macro setting. This paper primarily focuses on filling this gap.

This paper proposes a new framework for LP-IV models, which explicitly allows a LATE-like interpretation of IV estimand. The novelty is to observe that macro variables are often (if not always) aggregates of individuals or sectors whose dynamics may be heterogeneous. For example, consider the estimation of the government spending multiplier. Government spending is the sum of sectoral spendings such as defense and non-defense sectors, and their dynamic causal effects on the gross domestic product (GDP) would be heterogeneous. Using the structural vector moving average model, we show that the LP-IV estimand can be interpreted as a weighted average of sectoral dynamic causal effects whose weight depends on the response of the endogenous variable to the IV. Our theoretical framework builds on Stock and Watson (2018), who provided a theoretical ground for LP-IV.

We also extend our finding to the threshold LP-IV model, where the threshold parameter is given. It is standard practice to apply the LP-IV method to split samples such as recessions and expansions, or the zero lower bound (ZLB) and non-ZLB, where the sample split is based on a continuous threshold variable such as the unemployment rate or the interest rate, e.g., Ramey and Zubairy (2018). Allowing for a possibility that the threshold parameter is not correctly specified, we show that the threshold IV estimand is not only a weighted average of heterogeneous sectoral dynamic causal effects but also a weighted average of state-dependent dynamic causal effects. The second layer of the averaging (over the states) disappears when the threshold parameter is correctly specified, which makes the interpretation of the estimand as well as the expression simpler.

Finally, we develop a new statistical test for state-dependency robust to heterogeneous sectoral dynamic causal effects, and provide a rigorous distribution theory of the test statistic. It is a nontrivial extension of Chernozhukov, Chetverikov, and Kato (2014) to dependent data. Our test is relevant because the conventional test comparing two IV estimates based on the split sample is not consistent under heterogeneous dynamic causal effects.

We list some relevant work in the literature. Jordà and Taylor (2015), Ramey and Zubairy (2018), Jordà, Schularick, and Taylor (2019) are applications of LP-IV. Plagborg-Møller and Wolf (forthcoming) show that LP and VAR estimate the same population IRF's and the difference is due to their finite sample properties. They also show that LP-IV is equivalent to estimating a VAR with the IV ordered first in the system. Angrist, Jordà, and Kuersteiner (2018) provide the potential outcome framework in the time-series setting, but they do not discuss LATE.

2 LP-IV Estimator

For $t = 1, 2, ..., T$ and $h = 0, 1, 2, ..., H$, the local projection instrumental variable (LP-IV) model is given by

$$y_{t+h} = x_t'\beta_h + u_{t+h}, \tag{1}$$

where y_t is a scalar, x_t is $k \times 1$, and z_t is $l \times 1$. The instruments vector z_t satisfies $E[z_t u_{t+h}] = 0$ and $E[z_t x_t'] \neq 0$. If the model is just-identified ($l = k$), the IV estimator is

$$\widehat{\beta}_h = \left(\sum_{t=1}^{T} z_t x_t'\right)^{-1} \sum_{t=1}^{T} z_t y_{t+h} \tag{2}$$

and the IRFs are obtained by estimating $\widehat{\beta}_h$ for each $h = 0, 1, ..., H$. For over-identified models ($l > k$), the two-stage least squares (2SLS) estimator is

$$\widehat{\beta}_h = \left(\sum_{t=1}^{T} z_t x_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} x_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t x_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t y_{t+h}. \tag{3}$$

By replacing $\sum_{t=1}^{T} z_t z_t'$ with a general positive definite weight matrix, a GMM estimator may be obtained.

Now consider the threshold LP-IV model with state-dependent slope parameters:

$$y_{t+h} = x_t' 1(q_{t-1} \leq \gamma)\beta_{A,h} + x_t' 1(q_{t-1} > \gamma)\beta_{B,h} + u_{t+h}, \tag{4}$$

for $h = 0, 1, 2, ..., H$ and $t = 1, ..., T$, where q_t is the threshold variable. The instruments are $z_t 1(q_{t-1} \leq \gamma)$ and $z_t 1(q_{t-1} > \gamma)$. Here γ and $(\beta_{A,h}, \beta_{B,h})$ are unknown parameters but we assume that γ is given. Estimation of the threshold parameter in the IV regression is investigated in Caner and Hansen (2004) for an exogenous threshold variable. Seo and Shin (2016) extend the result for an endogenous threshold variable.

The IV estimators of the slope parameters given the threshold parameter for just-identified models are

$$\widehat{\beta}_{A,h}(\gamma) = \left(\sum_{t=1}^{T} z_t 1(q_{t-1} \leq \gamma) x_t'\right)^{-1} \sum_{t=1}^{T} z_t 1(q_{t-1} \leq \gamma) y_{t+h}, \tag{5}$$

$$\widehat{\beta}_{B,h}(\gamma) = \left(\sum_{t=1}^{T} z_t 1(q_{t-1} > \gamma) x_t'\right)^{-1} \sum_{t=1}^{T} z_t 1(q_{t-1} > \gamma) y_{t+h}. \tag{6}$$

The 2SLS estimators for over-identified models are defined similarly:

$$\widehat{\beta}_{A,h}(\gamma) = \left(\sum_{t=1}^{T} x_t z_t 1(q_{t-1} \leq \gamma)' \left(\sum_{t=1}^{T} z_t z_t' 1(q_{t-1} \leq \gamma)\right)^{-1} \sum_{t=1}^{T} z_t 1(q_{t-1} \leq \gamma) x_t'\right)^{-1} \tag{7}$$

$$\times \left(\sum_{t=1}^{T} x_t z_t 1(q_{t-1} \leq \gamma)' \left(\sum_{t=1}^{T} z_t z_t' 1(q_{t-1} \leq \gamma)\right)^{-1} \sum_{t=1}^{T} z_t 1(q_{t-1} \leq \gamma) y_{t+h}\right),$$

$$\widehat{\beta}_{B,h}(\gamma) = \left(\sum_{t=1}^{T} x_t z_t 1(q_{t-1} > \gamma)' \left(\sum_{t=1}^{T} z_t z_t' 1(q_{t-1} > \gamma)\right)^{-1} \sum_{t=1}^{T} z_t 1(q_{t-1} > \gamma) x_t'\right)^{-1} \tag{8}$$

$$\times \left(\sum_{t=1}^{T} x_t z_t 1(q_{t-1} > \gamma)' \left(\sum_{t=1}^{T} z_t z_t' 1(q_{t-1} > \gamma)\right)^{-1} \sum_{t=1}^{T} z_t 1(q_{t-1} > \gamma) y_{t+h}\right).$$

These estimators are equivalent to the IV estimator (2) applied to the split sample where the sample is split based on a continuous threshold parameter.

3 Identification using Structural Vector Moving Average Model

The IV estimand is the population version of the IV estimator. To characterize the IV estimand under heterogeneous dynamic causal effects, we consider the structural vector moving average (SVMA) model, which is also used in Stock and Watson (2018) and Plagborg-Møller and Wolf (forthcoming, 2020).

3.1 Baseline model under heterogeneous dynamic causal effects

For simplicity, assume that the model (1) has no control variables and x_t and z_t are scalars. An extension to models with control variables is straightforward by re-defining the variables as the projection residuals on the control variables. The baseline SVMA model can be written as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \Theta(L) \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix} \tag{9}$$

where L is the lag operator, $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \cdots$, and Θ_h for $h = 0, 1, 2, ...$ are 2×2 parameter matrices. Following Stock and Watson (2018), we normalize the diagonal elements of Θ_0 to be 1 (unit effect normalizations). The dynamic causal effect of $\varepsilon_{x,t}$ on y_{t+h} is

$$E[y_{t+h}|\varepsilon_{x,t} = 1] - E[y_{t+h}|\varepsilon_{x,t} = 0] = \theta_{h,yx} \tag{10}$$

where $\theta_{h,yx}$ is the $(1,2)$th element of the matrix Θ_h. Using the baseline model, Stock and Watson (2018) show that LP-IV estimand is $\theta_{h,yx}$.

The baseline model treats x_t and $\varepsilon_{x,t}$ as individual units, but typically they are aggregate by construction. For example, in Ramey and Zubairy (2018), government spending x_t is defined as the sum of all federal, state, and local purchases excluding transfer payments. Purchases can also be divided into sectors where the consumption is made, such as military, infrastructure, health, etc. Since a dollar spent on the military would have a quite different dynamic causal effect on the GDP compared to a dollar spent on health, it is reasonable to model this in the SVMA model explicitly.

We define $x_t = \sum_{s=1}^{S} x_{s,t}$ where $x_{s,t}$ is the sectoral component of the aggregate variable x_t. Let $\varepsilon_{s,t}$ be the structural shock to the sector s at time t. We consider the following SVMA model:

$$Y_t = \Theta(L)\varepsilon_t \tag{11}$$

where $Y_t = (y_t, x_{1,t}, x_{2,t}, ..., x_{S,t})'$, $\varepsilon_t = (\varepsilon_{y,t}, \varepsilon_{1,t}, \varepsilon_{2,t}, ..., \varepsilon_{S,t})'$, $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \cdots$, and Θ_h is $(S+1) \times (S+1)$ matrix for $h = 0, 1, 2, \cdots$. Similar to the baseline model (9), we assume the unit effect normalization, i.e., the diagonal elements of Θ_0 are equal to one. The parameter matrix Θ_h has the following structure:

$$\Theta_h = \begin{pmatrix} \theta_{h,yy} & \theta_{h,y1} & \cdots & \theta_{h,yS} \\ \theta_{h,1y} & \theta_{h,11} & \cdots & \theta_{h,1S} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{h,Sy} & \theta_{h,S1} & \cdots & \theta_{h,SS} \end{pmatrix} \tag{12}$$

where $\theta_{h,ys}$ and $\theta_{h,s's}$ are the dynamic causal effect of $\varepsilon_{s,t}$ on y_{t+h} and $x_{s',t+h}$, respectively. Similarly, $\theta_{h,sy}$ is defined as the dynamic causal effect of $\varepsilon_{y,t}$ on $x_{s,t+h}$. The dynamic causal effect of $\varepsilon_{s,t}$ on the sectoral sum x_{t+h} is compactly written as $\theta_{h,xs} = \sum_{s'=1}^{S} \theta_{h,s's}$. We assume that $\theta_{0,xs} \neq 0$ for all s. Since $\theta_{0,ss} = 1$ by the unit effect normalization, this assumption rules out the special case of $\sum_{s' \neq s} \theta_{0,s's} = -1$.

We make the following assumptions for the instrument z_t:

Assumption 1.

(i) $E[z_t \varepsilon_{s,t}] = \alpha_s \neq 0$ *for some* s *(relevance)*

(ii) $E[z_t \varepsilon_{y,t}] = 0$ *(contemporaneous exogeneity)*

(iii) $E[z_t \varepsilon_{t+j}] = 0$ *for* $j \neq 0$ *(lead-lag exogeneity)*

Assumption 1 is an extension of Condition LP-IV of Stock and Watson (2018) to heterogeneous sectors. Assumption 1(i) allows for heterogeneity in the response of sectors to the instrument. Assumption 1(ii)-(iii) are identical to Condition LP-IV (ii)-(iii) of Stock and Watson (2018).

Proposition 1. *Under Assumption 1, the IV estimand* β_h *for* $h = 0, 1, 2, ..., H$ *is given by*

$$\beta_h = \frac{E[z_t y_{t+h}]}{E[z_t x_t]} = \sum_{s=1}^{S} \left(\frac{\alpha_s \theta_{0,xs}}{\sum_{s'=1}^{S} \alpha_{s'} \theta_{0,xs'}} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}}.$$

Proposition 1 shows that the estimand β_h is a weighted average of the normalized sectoral dynamic causal effects $\theta_{h,ys}/\theta_{0,xs}$. The weights are $\alpha_s\theta_{0,xs}/\sum_{s'=1}^{S}\alpha_{s'}\theta_{0,xs'}$, which depend on α_s measuring the response of sector s to the instrument z_t, and the contemporaneous causal effect of $\varepsilon_{s,t}$ on the sectoral sum x_t. The weights can be negative but sum to one. If we further assume that the instrument z_t only affects sector s, i.e., $\alpha_{s'} = 0$ for $s' \neq s$, then

$$\beta_h = \frac{\theta_{h,ys}}{\theta_{0,xs}}, \tag{13}$$

which is the dynamic causal effect of sector s.

The LP-IV estimand can be interpreted as the local average treatment effect (LATE) of Imbens and Angrist (1994). LATE is the average treatment effect of the subpopulation called compliers who receive treatment because of the instrument. Since the complier subpopulation is defined by the instrument, LATE is instrument-specific. The LP-IV estimand has similar properties. First, it is a weighted average of the sectoral dynamic causal effects with larger weights on sectors that respond stronger to the instrument. Second, it depends on the instrument. On the other hand, sectors are known and observed while compliers are not identified individually.[1]

Even with observed sectors, the weights in the IV estimand in Proposition 1 are not identified in general without further restrictions. Proposition 2 gives a condition under which the sector-specific weights are identified.

Proposition 2. *Under Assumption 1 and no inter-sectoral causal effects, $\theta_{0,s's} = 0$ for all $s' \neq s$, the IV estimand β_h for $h = 0, 1, 2, ..., H$ can be written as*

$$\beta_h = \frac{E[z_t y_{t+h}]}{E[z_t x_t]} = \sum_{s=1}^{S} \frac{E[z_t x_{s,t}]}{E[z_t x_t]}\theta_{h,ys}.$$

Proposition 2 shows that the sector-specific weights are identified (i.e., can be written as a function of data moments) under no inter-sectoral causal effects. This result may be useful for counterfactual analyses. Suppose $S = 2$ for simplicity. The parameters of interest are $\theta_{h,y1}$ and $\theta_{h,y2}$. Since β_h and the weights are consistently estimated by the sample moments, we can construct a confidence interval for $\theta_{h,y1}$ for a given $\theta_{h,y2}$ and vice versa. Alternatively, if an additional instrument is available, then both of $\theta_{h,y1}$ and $\theta_{h,y2}$ can be identified. In principle, $\theta_{h,ys}$ for $s = 1, ..., S$ are identified if S instruments are available.

The zero inter-sectoral causal effects assumption is critical for identification of the sectoral components. Without the assumption, the ratio $E[z_t x_{s,t}]/E[z_t x_t]$ does not identify the weight. Moreover, the IV estimand using z_t for the sectoral component $x_{s,t}$, $E[z_t y_{t+h}]/E[z_t x_{s,t}]$, is still a weighted average of the sectoral dynamic causal effects, rather than the sector-specific dynamic causal effect.

[1] Only their statistical characteristics are identified, see Abadie (2003).

3.2 Threshold model under heterogeneous dynamic causal effects

We extend the SVMA model (11) under sectoral heterogeneity to threshold models. The threshold SVMA model is given by

$$Y_t = (\Theta(L)1(q_{t-1} \leq \gamma_0) + \Psi(L)1(q_{t-1} > \gamma_0))\, \varepsilon_t, \tag{14}$$

where $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \cdots$, $\Psi(L) = \Psi_0 + \Psi_1 L + \Psi_2 L^2 + \cdots$, and q_t is the threshold variable. We assume that q_{t-1} is externally determined so that q_t does not appear on the left-hand side of (14). The diagonal elements of Θ_0 and Ψ_0 are normalized to one (the unit effect normalization). The matrices Θ_h and Ψ_h are similarly defined with (12). We write the state-dependent dynamic causal effect of $\varepsilon_{s,t}$ on x_t as $\theta_{h,xs} = \sum_{s'=1}^{S} \theta_{h,s's}$ and $\psi_{h,xs} = \sum_{s'=1}^{S} \psi_{h,s's}$, respectively. We also assume that $\theta_{0,xs} \neq 0$ and $\psi_{0,xs} \neq 0$ for all s, which are mild under the unit effect normalization.

We consider the IV estimators using $z_t 1(q_{t-1} \leq \gamma)$ and $z_t 1(q_{t-1} > \gamma)$ as instruments, allowing for $\gamma \neq \gamma_0$. Define

$$\begin{aligned}
\alpha_{A,s}(\gamma) =& E[z_t \varepsilon_{s,t} 1(q_{t-1} \leq \gamma)], \\
\alpha_{B,s}(\gamma) =& E[z_t \varepsilon_{s,t} 1(q_{t-1} > \gamma)], \\
\alpha_s(\gamma_1, \gamma_2) =& E[z_t \varepsilon_{s,t} 1(\gamma_1 < q_{t-1} \leq \gamma_2)] \quad \text{for } \gamma_1 \leq \gamma_2.
\end{aligned}$$

By construction, $\alpha_s = \alpha_{A,s}(\gamma) + \alpha_{B,s}(\gamma)$. The instruments satisfy the following assumptions:

Assumption 2.

(i) $\alpha_{A,s}(\gamma) \neq 0$ *and* $\alpha_{B,s}(\gamma) \neq 0$, $\forall \gamma \in \Gamma$, *for some* s *(relevance)*

(ii) $E[z_t \varepsilon_{y,t} | q_{t-1}] = 0$ *(contemporaneous exogeneity)*

(iii) For $j \geq 1$, $E[z_t \varepsilon_{t+j} | q_{t-1}] = 0$ *and* $E[z_t \varepsilon_{t-j} | q_{t-1}] = 0$ *(lead-lag exogeneity)*

Assumption 2(i) is the relevance condition for the instrument of the split sample. Since $\lim_{\gamma \to -\infty} \alpha_{A,s}(\gamma) = \lim_{\gamma \to \infty} \alpha_{B,s}(\gamma) = 0$, it restricts the support of Γ, the parameter space of γ. Assumption 2(ii) is the exclusion restriction conditional on the threshold variable. Assumption 2(iii) is the threshold version of Assumption 1(iii), stating that z_t is uncorrelated with all future and past ε's conditional on the threshold variable. Assumptions 2(ii)-(iii) are in the form of conditional expectations, which are slightly stronger than necessary but give simpler results.

Let the split-sample IV estimands be

$$\beta_{A,h}(\gamma) = \frac{E[z_t 1(q_{t-1} \leq \gamma) y_{t+h}]}{E[z_t 1(q_{t-1} \leq \gamma) x_t]}, \tag{15}$$

$$\beta_{B,h}(\gamma) = \frac{E[z_t 1(q_{t-1} > \gamma) y_{t+h}]}{E[z_t 1(q_{t-1} > \gamma) x_t]}, \tag{16}$$

for $h = 0, 1, 2, ..., H$. The following proposition characterizes the IV estimands under sectoral heterogeneity.

Proposition 3. *Under Assumption 2, the IV estimands $\beta_{A,h}(\gamma)$ and $\beta_{B,h}(\gamma)$ for $h = 0, 1, 2, ..., H$ are given by as follows. If $\gamma \leq \gamma_0$,*

$$\beta_{A,h}(\gamma) = \sum_{s=1}^{S} \left(\frac{\alpha_{A,s}(\gamma)\theta_{0,xs}}{\sum_{s'=1}^{S} \alpha_{A,s'}(\gamma)\theta_{0,xs'}} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}},$$

$$\beta_{B,h}(\gamma) = \sum_{s=1}^{S} \left[\left(\frac{\alpha_{B,s}(\gamma_0)\psi_{0,xs}}{\sum_{s'=1}^{S}(\alpha_{B,s'}(\gamma_0)\psi_{0,xs'} + \alpha_{s'}(\gamma,\gamma_0)\theta_{0,xs'})} \right) \frac{\psi_{h,ys}}{\psi_{0,xs}} + \left(\frac{\alpha_s(\gamma,\gamma_0)\theta_{0,xs}}{\sum_{s'=1}^{S}(\alpha_{B,s'}(\gamma_0)\psi_{0,xs'} + \alpha_{s'}(\gamma,\gamma_0)\theta_{0,xs'})} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}} \right],$$

and if $\gamma > \gamma_0$,

$$\beta_{A,h}(\gamma) = \sum_{s=1}^{S} \left[\left(\frac{\alpha_{A,s}(\gamma_0)\theta_{0,xs}}{\sum_{s'=1}^{S}(\alpha_{A,s'}(\gamma_0)\theta_{0,xs'} + \alpha_{s'}(\gamma_0,\gamma)\psi_{0,xs'})} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}} + \left(\frac{\alpha_s(\gamma_0,\gamma)\psi_{0,xs}}{\sum_{s'=1}^{S}(\alpha_{A,s'}(\gamma_0)\theta_{0,xs'} + \alpha_{s'}(\gamma_0,\gamma)\psi_{0,xs'})} \right) \frac{\psi_{h,ys}}{\psi_{0,xs}} \right],$$

$$\beta_{B,h}(\gamma) = \sum_{s=1}^{S} \left(\frac{\alpha_{B,s}(\gamma)\psi_{0,xs}}{\sum_{s'=1}^{S} \alpha_{B,s'}(\gamma)\psi_{0,xs'}} \right) \frac{\psi_{h,ys}}{\psi_{0,xs}}.$$

Similar to Proposition 1, Proposition 3 shows that the estimand of the split-sample IV estimator is a weighted average of the normalized state-dependent sectoral dynamic causal effects. The estimand is averaged across the states due to potential misspecification of the threshold parameter.

The IV estimands in Proposition 3 simplify if the threshold parameter is correctly specified. Setting $\gamma = \gamma_0$, we can write

$$\beta_{A,h}(\gamma) = \sum_{s=1}^{S} \left(\frac{\alpha_{A,s}(\gamma_0)\theta_{0,xs}}{\sum_{s'=1}^{S} \alpha_{A,s'}(\gamma_0)\theta_{0,xs'}} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}}, \tag{17}$$

$$\beta_{B,h}(\gamma) = \sum_{s=1}^{S} \left(\frac{\alpha_{B,s}(\gamma_0)\psi_{0,xs}}{\sum_{s'=1}^{S} \alpha_{B,s'}(\gamma_0)\psi_{0,xs'}} \right) \frac{\psi_{h,ys}}{\psi_{0,xs}} \tag{18}$$

If we further assume $\alpha_{A,s'}(\gamma_0) = \alpha_{B,s'}(\gamma_0) = 0$ for all $s' \neq s$, we can write

$$\beta_{A,h}(\gamma_0) = \frac{\theta_{h,ys}}{\theta_{0,xs}}, \tag{19}$$

$$\beta_{B,h}(\gamma_0) = \frac{\psi_{h,ys}}{\psi_{0,xs}}. \tag{20}$$

Proposition 4 shows that the weights in the IV estimands are identified if inter-sectoral causal effects are zero and the threshold parameter is correctly specified.

Proposition 4. *Under Assumption 2, no inter-sectoral causal effects, $\theta_{0,s's} = \psi_{0,s's} = 0$ for all*

$s' \neq s$, and the correct specification of the threshold parameter, $\gamma = \gamma_0$, the IV estimands $\beta_{A,h}$ and $\beta_{B,h}$ for $h = 0, 1, 2, ..., H$ can be written as

$$\beta_{A,h}(\gamma_0) = \sum_{s=1}^{S} \frac{E[z_t 1(q_{t-1} \leq \gamma_0) x_{s,t}]}{E[z_t 1(q_{t-1} \leq \gamma_0) x_t]} \theta_{h,ys},$$
$$\beta_{B,h}(\gamma_0) = \sum_{s=1}^{S} \frac{E[z_t 1(q_{t-1} > \gamma_0) x_{s,t}]}{E[z_t 1(q_{t-1} > \gamma_0) x_t]} \psi_{h,ys}.$$

4 A New Test for State-Dependency

In this section, we show that the conventional test for state-dependency is invalid in the presence of heterogeneous sectoral dynamic causal effects and propose a new test for state-dependency robust to such heterogeneity.

4.1 Conventional test for state-dependency

The standard approach to testing state-dependency of the slope parameter is to split the sample with respect to the assumed continuous threshold parameter and compare the estimated values based on each of the split sample. For a given threshold γ^*, the null hypothesis of the conventional test is

$$H_0 : \beta_{A,h}(\gamma^*) = \beta_{B,h}(\gamma^*), \tag{21}$$

where $\beta_{A,h}(\gamma)$ and $\beta_{B,h}(\gamma)$ are given by (15) and (16). The conventional test statistic is the standardized value of $|\widehat{\beta}_{A,h}(\gamma^*) - \widehat{\beta}_{B,h}(\gamma^*)|$. This test has been widely used in practice either formally or informally, see Ramey and Zubairy (2018), for instance. We show this conventional approach is not appropriate under heterogeneous sectoral dynamic causal effects.

Assume that the threshold parameter is correctly specified, $\gamma^* = \gamma_0$, and no state-dependency in the dynamic causal effect, i.e., $\theta_{h,ys} = \psi_{h,ys}$ and $\theta_{0,xs} = \psi_{0,xs}$ for all s. The population version of the test statistic should be zero under the null hypothesis. However, using Proposition 3, the IV estimands can be written as

$$\beta_{A,h}(\gamma_0) = \sum_{s=1}^{S} \left(\frac{\alpha_{A,s}(\gamma_0)\theta_{0,xs}}{\sum_{s'=1}^{S} \alpha_{A,s'}(\gamma_0)\theta_{0,xs'}} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}},$$
$$\beta_{B,h}(\gamma_0) = \sum_{s=1}^{S} \left(\frac{\alpha_{B,s}(\gamma_0)\theta_{0,xs}}{\sum_{s'=1}^{S} \alpha_{B,s'}(\gamma_0)\theta_{0,xs'}} \right) \frac{\theta_{h,ys}}{\theta_{0,xs}}.$$

Thus, $\beta_{A,h}(\gamma_0) = \beta_{B,h}(\gamma_0)$ if $\alpha_{A,s}(\gamma_0) = \alpha_{B,s}(\gamma_0)$ for all s, or equivalently,

$$E[z_t \varepsilon_{s,t} 1(q_{t-1} \leq \gamma_0)] = E[z_t \varepsilon_{s,t} 1(q_{t-1} > \gamma_0)], \ \forall s. \tag{22}$$

But this may not be true in general. In particular, under the assumption of Proposition 4, the condition (22) can be replaced with

$$E[z_t x_{s,t} 1(q_{t-1} \leq \gamma_0)] = E[z_t x_{s,t} 1(q_{t-1} > \gamma_0)], \ \forall s, \tag{23}$$

which further implies

$$E[z_t x_t 1(q_{t-1} \leq \gamma_0)] = E[z_t x_t 1(q_{t-1} > \gamma_0)]. \tag{24}$$

Therefore, if (24) is violated then we may incorrectly reject the null hypothesis of no state-dependency. Both of (23) and (24) are testable and they can be conveniently verified in practice.

4.2 New test for state-dependency

Our new test builds on the fact that the IV estimand is a kinked function of the threshold parameter with the kink at $\gamma = \gamma_0$ in the presence of state-dependency using the threshold SVMA model (14). We first assume the following regularity conditions.

Assumption 3.

(i) $\gamma_0 \in int(\Gamma)$

(i) $\alpha_{A,s}(\gamma)$ *and* $\alpha_{B,s}(\gamma)$ *are differentiable on* $int(\Gamma)$

(ii) The probability dentity function of q *at* $q = \gamma$, $p_q(\gamma)$, *is smooth over* Γ

Consider $\beta_{A,h}(\gamma)$ in (15). Under Assumption 2, its numerator is

$$E[z_t 1(q_{t-1} \leq \gamma) y_{t+h}] = \begin{cases} \sum_{s=1}^{S} \alpha_{A,s}(\gamma)\theta_{h,ys} & \text{if } \gamma \leq \gamma_0, \\ \sum_{s=1}^{S} (\alpha_{A,s}(\gamma_0)\theta_{h,ys} + \alpha_s(\gamma_0, \gamma)\psi_{h,ys}) & \text{if } \gamma > \gamma_0. \end{cases} \tag{25}$$

The left and right derivatives of (25) around $\gamma = \gamma_0$ are

$$\lim_{\epsilon \downarrow 0} \frac{E[z_t 1(q_{t-1} \leq \gamma_0) y_{t+h}] - E[z_t 1(q_{t-1} \leq \gamma_0 - \epsilon) y_{t+h}]}{\epsilon} = \sum_{s=1}^{S} \alpha'_{A,s}(\gamma_0)\theta_{h,ys}, \tag{26}$$

$$\lim_{\epsilon \downarrow 0} \frac{E[z_t 1(q_{t-1} \leq \gamma_0 + \epsilon) y_{t+h}] - E[z_t 1(q_{t-1} \leq \gamma_0) y_{t+h}]}{\epsilon} = \sum_{s=1}^{S} \alpha'_{A,s}(\gamma_0)\psi_{h,ys}, \tag{27}$$

where $\alpha'_{A,s}(\gamma) = (\partial/\partial\gamma)\alpha_{A,s}(\gamma)$. Thus, as long as $\theta_{h,ys} \neq \psi_{h,ys}$ for some s such that $\alpha'_{A,s}(\gamma_0) \neq 0$, $E[z_t 1(q_{t-1} \leq \gamma) y_{t+h}]$ exhibits a kink at $\gamma = \gamma_0$.

Since we can write by the fundamental theorem of calculus

$$\frac{\partial}{\partial\gamma} E[z_t 1(q_t \leq \gamma) y_{t+h}] = E[z_t y_{t+h} | q_{t-1} = \gamma] p_q(\gamma), \tag{28}$$

$$\frac{\partial}{\partial\gamma} E[z_t 1(q_t > \gamma) y_{t+h}] = -E[z_t y_{t+h} | q_{t-1} = \gamma] p_q(\gamma), \tag{29}$$

the derivatives will exhibit a jump (discontinuity) at $\gamma = \gamma_0$ under Assumption 3. The test for state-dependency is therefore equivalent to a test for the presence of a jump:

$$H_0 : \ E[z_t y_{t+h} | q_{t-1} = \gamma] \text{ has no jump.} \tag{30}$$

For the null hypothesis of no state-dependency in (30), the natural test statistic is the supremum of appropriately scaled differences between estimates of the left and right limits of the function $E[z_t y_{t+h} | q_{t-1} = \gamma]$ over the interior of Γ. The test statistic is constructed using a nonparametric one-sided kernel estimation approach.

More specifically, consider the following regression model,

$$g_t = \mu(q_{t-1}) + u_t \tag{31}$$

where g_t is the multiple of an instrument, z_t and a dependent variable, y_{t+h}, $q_{t-1} \in \Gamma$ is a predetermined state variable, and $\mu(\gamma) = E[g_t | q_{t-1} = \gamma]$ by construction.

Under the setup, we want to see whether there is any discontinuity in $\mu(\cdot)$ without specifying its functional form. That is, we want to test

$$\begin{cases} H_0 : & \mu(\cdot) \text{ is a continuous function on the domain of the threshold variable, } q_{t-1}. \\ H_1 : & \text{At least one jump exists.} \end{cases} \tag{32}$$

We let $\mu(\cdot)$ be a nonparametric function given the complicated and approximating nature of the quantity. This also reflects the fact that the local projection is only an approximation, and hence it is better for its form to be unspecified except for some minimal regularity conditions.

We construct a test statistic based on the difference between the right and left limits of the regression function via the one-side kernel estimation. More specifically, define $\mu^+(\gamma) = \lim_{x \downarrow \gamma} \mu(x)$ and $\mu^-(\gamma) = \lim_{x \uparrow \gamma} \mu(x)$ to denote the right and left limits of the function $\mu(\cdot)$. We assume that if there exists a jump, the jump location is an interior point in the domain Γ of the threshold variable.

The test statistic we consider is a sup-type one defined as

$$J_T = \sup_{\gamma \in \Gamma} \left\{ \sqrt{\frac{Tb}{\hat{\nu}(\gamma)}} \left| \hat{\mu}^+(\gamma) - \hat{\mu}^-(\gamma) \right| \right\} \tag{33}$$

where

$$\hat{\mu}^{\pm}(\gamma) = \frac{\sum_t^T K_b^{\pm}(\gamma - q_{t-1}) g_t}{\sum_t^T K_b^{\pm}(\gamma - q_{t-1})}$$

with $K_b^{\pm}(\cdot) = K^{\pm}(\cdot / b)$ and a scaling factor $\nu(\gamma)$ obtained by

$$\hat{\nu}(\gamma) = \frac{\hat{\sigma}^{2+}(\gamma)}{\hat{p}^+(\gamma)} + \frac{\hat{\sigma}^{2-}(\gamma)}{\hat{p}^-(\gamma)}$$

with

$$\hat{\sigma}^{2\pm}(\gamma) = \frac{\sum K_b^{\pm}(\gamma - q_{t-1})(g_t - \hat{\mu}(q_{t-1}))^2}{\sum_{t=1} K_b^{\pm}(\gamma - q_{t-1})} \quad \text{and} \quad \hat{p}^{\pm}(\gamma) = \sum K_b^{\pm}(\gamma - q_{t-1}).$$

The novelty of the test is that state-dependency of the parameter of interest is tested without relying on consistent estimation of coefficients and knowledge of correct specification of models and heterogeneity. The test is therefore nonparametric by nature, and it is robust to heterogeneity in the sectoral dynamic causal effects.

4.3 Distribution theory for testing the state-dependency

The test statistic given in (33) is a supremum-type based on nonparametric kernel estimates, and therefore deriving its null distribution is nonstandard. It is nontrivial to derive the distribution of the test statistic as in Theorem 1 especially in the context of time series. Therefore, it is of independent interest and a significant econometric and statistical contribution to the literature since existing methods cannot be directly implemented. Utilizing a recent development of theories on empirical processes, the theory entails approximating suprema of local empirical processes with dependent data by suprema of Gaussian processes (strong approximation). More technical details and regularity conditions regarding strong approximation is provided in Appendix B. We present an approximating distribution of the test statistic in the form of Theorem 1.

Theorem 1. *Under Assumption 4 in Appendix B, there exists a random variable $\mathcal{J}_T \overset{d}{\sim} \sup_{f \in \mathcal{F}_T} G_T f$ where G_T is a tight Gaussian process indexed by $\mathcal{F}_T$ where*

$$\mathcal{F}_T = \left\{ f_q(x,y) : \left(K_b^+(q-x) - K_b^-(q-x)\right) y/\sqrt{b} \text{ where } q \in \Gamma \right\}.$$

and covariance function

$$E G_T(g) G_T(f) = \frac{1}{\nu} Cov\left(g\left(\tilde{\xi}_t\right), f\left(\tilde{\xi}_t\right)\right)$$

where $\{\tilde{\xi}_t\}$ is an MDS defined in Lemma 2 and $g, f \in \mathcal{F}_T$ and $\nu = 2\sigma^2(q)/f(q)$ and we have

$$|J_T - \mathcal{J}_T| = O_p\left(A_T + B_T + C_T\right),$$

$$\begin{aligned} A_T &= m^{1/2-1/q} b^{-1/2} T^{-1/2+1/q} \log^{3/2} T \\ B_T &= m^{1/4} (bT)^{-1/4} \log^{5/4} T \\ C_T &= (bT)^{-1/6} \log T \end{aligned}$$

where m is the size of each big block in the Bernstein's big-and-small decomposition in Appendix B and b is the bandwidth.

Theorem 1 entails (i) approximating the original data by big and small independent blocks whose size are m and r respectively and (ii) approximating the big independent blocks by martingale difference sequences (MDS) while the small independent blocks are negligible in the limit. This work is a considerable time-series extension of recent theoretical development of empirical processes in Chernozhukov, Chetverikov and Kato (2014), which is only concerned with independent and identically distributed (i.i.d.) data.

Our result is also a nontrivial generalization of Antoch, Gregoire and Hukov (2007) who investigated the asymptotic behavior of a test statistic for continuity of regression functions based on i.i.d. data.

The approximating distribution can be simulated using the plug-in method, and critical values with selected significant levels can be tabulated for inference. However, bootstrap techniques are preferred in practice, given that the covariance function contains nuisance parameters for both computational convenience and precision.

4.3.1 Construction of null distribution: Residual wild bootstrapping

Since the convergence to the approximating distribution discussed in Theorem 1 is rather slow, and the asymptotic distribution involves nuisance parameters, the bootstrap method is employed in practice. Due to the lack of serial correlation structure of approximating distribution in Theorem 1, we employ the residual wild bootstrap (RWB) method.

The algorithm generating paired bootstrap data $\{(g_t^{*,(j)}, q_{t-1}^{*,(j)})\}_{t=1}^T; j = 1, ..., B$ from the paired observations $\{(g_t, q_{t-1})\}_{t=1}^T$ is given in Algorithm 1.

Algorithm 1 Construction of Null Distribution via RWB

1: Given (31), we estimate $\widehat{\mu}(q)$ under the null via the two sided kernel estimation:

$$\widehat{\mu}(q) = \frac{\sum_t^T K_b(q - q_{t-1})g_t}{\sum_t^T K_b(q - q_{t-1})}$$

where $K_b(\cdot)$ is a symmetric kernel function.

2: Obtain residuals from $\widehat{\mu}(q)$:

$$\widehat{u}_t = g_t - \widehat{\mu}(q_{t-1}).$$

3: Construct the paired bootstrap data $\{(g_t^{*,(j)}, q_{t-1}^{*,(j)})\}_{t=1}^T$:

$$g_t^{*,(j)} = \widehat{\mu}(q_{t-1}^{*,(j)}) + w_t \widehat{u}_t$$

where $q_{t-1}^{*,(j)} = q_{t-1}$ and w_t is an i.i.d. random variable with mean zero and variance one.

4: Repeat the procedure given the desired number of bootstrap repetitions.

After B repetitions, construct the empirical distribution of bootstrapped test Statistics $\mathcal{T}_T^*$ under the null. Note that the validity of the suggested bootstrapping is studied in Antoch et al. (2007).

5 Non-military spending multiplier in the U.S.

The government spending multiplier is the ratio of the change in the GDP to the change in government spending. The magnitude of the multiplier is the key to making fiscal policies, but the academic community has not been able to reach a consensus on it. Different models, shocks, and sample periods often give quite different estimates of the multiplier, e.g., Blanchard and Perotti (2002), Barro and Redlick (2011), Ramey (2011), Auerbach and Gorodnichenko (2012), Nakamura and Steinson (2014), and Ramey and Zubairy (2018), among many important others.

In this section, we estimate the non-military spending multiplier in the U.S. by applying our method to Ramey and Zubairy (2018), who estimate the cumulative government spending multiplier using the military news shock as the instrument. We first replicate their results to obtain the cumulative government spending multiplier estimate, which is denoted by $\widehat{m}_h$. Measured as the two-year or four-year integral, we find that $\widehat{m}_h < 1$, which is consistent with Ramey and Zubairy (2018). Then we decompose $\widehat{m}_h$ into a weighted average of military and non-military spending multipliers. The weights are estimated by the response of sectoral spending to the military news shock. The post-WWII data shows that the response of the non-military spending to the military news shock is negative and persistent, which leads to a negative weight for the non-military spending multiplier. As a result, we find that the less-than-unity aggregate spending multiplier can be written as a weighted average of larger-than-unity non-military and military spending multipliers.

The model in the linear case (no state-dependency) is

$$\sum_{j=0}^{h} y_{t+j} = c_h + \phi_h(L)w_{t-1} + m_h \sum_{j=0}^{h} g_{t+j} + \epsilon_{t+h}, \tag{34}$$

where y_t is the GDP, g_t is the government spending, w_t is a set of control variables, and $\phi_h(L)$ is a polynomial in the lag operator of order 4. Since $\sum_{j=0}^{h} y_{t+j}$ is the sum of the GDP over h periods and $\sum_{j=0}^{h} g_{t+j}$ is the sum of government spending over h periods, the parameter m_h is the cumulative government spending multiplier. Ramey and Zubairy (2018) argue that the cumulative multiplier is more appropriate in a dynamic setting than other definitions. Since g_t is endogenous, we use the military news shock as the IV, denoted by z_t. The control variables include lagged values of y_t, g_t, and z_t. The state-dependent case is given by

$$\begin{aligned}\sum_{j=0}^{h} y_{t+j} =& (c_{A,h} + \phi_{A,h}(L)w_{t-1} + m_{A,h} \sum_{j=0}^{h} g_{t+j})1(q_{t-1} \leq \gamma^*) \\ &+ (c_{B,h} + \phi_{B,h}(L)w_{t-1} + m_{B,h} \sum_{j=0}^{h} g_{t+j})1(q_{t-1} < \gamma^*) + \epsilon_{t+h},\end{aligned} \tag{35}$$

where q_t is the unemployment rate and γ^* is a given threshold. In both models, the parameters are estimated by the just-identified IV estimator.

Define the cumulative variables $\widetilde{y}^{\perp}_{t+h} = \sum_{j=0}^{h} y^{\perp}_{t+j}$ and $\widetilde{x}^{\perp}_{t+h} = \sum_{j=0}^{h} x^{\perp}_{t+j}$. By the Frisch-Waugh-

Lovell (FWL) theorem[2], the IV estimand for the cumulative multiplier in the linear is

$$m_h = \frac{E[z_t^{\perp}\widetilde{y}_{t+h}^{\perp}]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}]}, \tag{36}$$

where $y_t^{\perp}$, $x_t^{\perp}$, and $z_t^{\perp}$ as the residual of the original variable after projecting onto the control variables including the intercept in (34). In the state-dependent case,

$$m_{A,h}(\gamma^*) = \frac{E[z_t^{\perp}\widetilde{y}_{t+h}^{\perp}1(q_{t-1} \le \gamma^*)]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}1(q_{t-1} \le \gamma^*)]}, \tag{37}$$

$$m_{B,h}(\gamma^*) = \frac{E[z_t^{\perp}\widetilde{y}_{t+h}^{\perp}1(q_{t-1} > \gamma^*)]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}1(q_{t-1} > \gamma^*)]} \tag{38}$$

where $y_t^{\perp}$, $x_t^{\perp}$, and $z_t^{\perp}$ as the residual of the original variable after projecting onto the control variables including the intercept in (35). Note that the residuals $y_t^{\perp}$, $x_t^{\perp}$, and $z_t^{\perp}$ in (36) and in (38) are different as they are based on different equations. We abuse the notation to keep the expression simple.

Assume two sectors ($S = 2$): the military sector ($s = 1$) and the non-military sector ($s = 2$). Let $x_{1,t}$ and $x_{2,t}$ be military spending, and non-military spending, respectively. Based on the SVMA model (11) and Proposition 2, the cumulative multiplier can be decomposed as

$$m_h = \frac{E[z_t^{\perp}\widetilde{x}_{1,t+h}^{\perp}]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}]}\widetilde{\theta}_{h,y1} + \frac{E[z_t^{\perp}\widetilde{x}_{2,t+h}^{\perp}]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}]}\widetilde{\theta}_{h,y2}, \tag{39}$$

where $\widetilde{\theta}_{h,ys} = \sum_{j=0}^{h} \theta_{j,ys}$ for $s = 1, 2$. The cumulative dynamic causal effects $\widetilde{\theta}_{h,y1}$ and $\widetilde{\theta}_{h,y2}$ are the military and the non-military spending multipliers, respectively. Assuming correct specification of the threshold, $\gamma^* = \gamma_0$, the state-dependent cumulative multipliers can be similarly decomposed using Proposition 4 as

$$m_{A,h}(\gamma^*) = \frac{E[z_t^{\perp}\widetilde{x}_{1,t+h}^{\perp}1(q_{t-1} \le \gamma^*)]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}1(q_{t-1} \le \gamma^*)]}\widetilde{\theta}_{h,y1} + \frac{E[z_t^{\perp}\widetilde{x}_{2,t+h}^{\perp}1(q_{t-1} \le \gamma^*)]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}1(q_{t-1} \le \gamma^*)]}\widetilde{\theta}_{h,y2}, \tag{40}$$

$$m_{B,h}(\gamma^*) = \frac{E[z_t^{\perp}\widetilde{x}_{1,t+h}^{\perp}1(q_{t-1} > \gamma^*)]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}1(q_{t-1} > \gamma^*)]}\widetilde{\psi}_{h,y1} + \frac{E[z_t^{\perp}\widetilde{x}_{2,t+h}^{\perp}1(q_{t-1} > \gamma^*)]}{E[z_t^{\perp}\widetilde{x}_{t+h}^{\perp}1(q_{t-1} > \gamma^*)]}\widetilde{\psi}_{h,y2}, \tag{41}$$

where $\widetilde{\psi}_{h,ys} = \sum_{j=0}^{h} \psi_{j,ys}$ for $s = 1, 2$ is the state-dependent cumulative sectoral multiplier. Since the weights are the population moments of the observables, they can be consistently estimated by the sample moments.

We use the post-WWII data as the quarterly defense spending data is only available from 1947Q1. The econometric model specification and the control variables are identical to Ramey and Zubairy (2018). The federal defense consumption expenditures and gross investment data

[2] See p.344 of Hansen (2020) for the FWL theorem applied to 2SLS.

(FDEFX) is obtained from the Federal Reserve Bank of St. Louis FRED database. The sample period is from 1947Q1 to 2015Q4.

Figure 1 shows real government spending, defense spending, and non-defense spending per capita in the U.S. The vertical lines represent Ramey and Shapiro (1998) dates, including the Korean War, the Vietnam War, and the Soviet invasion of Afghanistan, and 9/11. The defense spending gradually increases after the major war events and decreases over time. The share of defense spending has been decreased over time.

Figure 2 shows the de-trended real series of government spending, defense spending, non-defense spending, and military news shock. The relative magnitude of the news shock before the Korean War is very large, so we provide a robustness check excluding the Korean War later in the section.

Figure 3 shows the cumulative government spending multiplier in the linear model (no state-dependency) for each horizon from two quarters to 5 years out. The bands are 95% confidence bands using the Newey-West standard errors (Newey and West, 1987). The top panel shows the cumulative spending multiplier, which corresponds to the top panel of Figure 6 of Ramey and Zubairy (2018). Note that ours is based on the sample from 1947Q1 to 2015Q4, while Figure 6 of Ramey and Zubairy (2018) is based on the sample from 1889Q1 to 2015Q4. The result is consistent with Ramey and Zubairy (2018) that the aggregate multiplier is less than one.

Our sectoral decomposition reveals interesting trajectories of the non-military spending multiplier. The bottom panel of Figure 3 shows the cumulative non-military spending multiplier when the military spending multiplier is calibrated at 1.2 (solid blue line) and at 0.6 (red circle marker). The non-military spending multiplier gradually increases over time in both cases. More importantly, it is possible that both the military and the non-military spending multipliers are higher than one while $\widehat{m}_h$ is less than one. In this case, the non-military spending multiplier is significantly higher than the military spending multiplier, which is set to 1.2.

This result is due to the negative weight on the non-military spending multiplier in the decomposition. Figure 4 shows the IRFs of total and sectoral spending to the news shock. The weights are given by the ratio of the sectoral IRF to that of total spending. The top panel shows the IRFs of total spending (black solid), military spending (blue dotted), and non-military spending (red circle marker). The IRF of non-military spending is negative (statistically significant) and persistent. This negative IRF leads to a negative weight for the non-military spending multiplier, as shown in Figure 5.

Next we investigate if the non-military spending multiplier is state-dependent. Auerbach and Gorodnichekko (2012) found strong evidence on the state-dependency of the multiplier using the smooth transition VAR, but Ramey and Zubairy (2018) found little evidence supporting the state-dependency using the LP-IV. Ramey and Zubairy (2018) use the unemployment rate of 6.5% as the threshold but this is not an estimated value. Since our decomposition result for the state-dependent case critically relies on the correct specification of the threshold parameter, we apply our new test for state-dependency.

Since 1947Q1, the lowest unemployment rate is 2.6% in 1953Q2 and the highest is 10.7%

in 1982Q4. Figure 6 shows the value of the test statistic against the unemployment rate. The bootstrap 95% critical value is 13.97, so we would reject the null hypothesis if the test statistic is greater than the critical value. Using all the observations from 1947Q1 to 2015Q4, the maximum of the test statistic is 6.73 at the unemployment rate of 5.517%. But this result is largely driven by one influential observation during the Korean War. When this observation (1950Q3) is removed, the maximum of 2.06. In both cases, we do not reject the null hypothesis of the existence of at least one jump (state-dependency) at the 5% level. We conclude that there is little evidence of state-dependency after WWII.

Although we do not find strong evidence for state-dependency, we provide additional analyses for state-dependent multipliers. We set $\gamma^* = 5.6$ so that observations with the unemployment rate equals to or lower than (higher than) 5.6% are defined as the low-employment (high-employment) state.

Figures 7 and 8 are the cumulative spending multipliers and the IRFs. Overall, the results are similar to the linear case: The non-military spending multiplier is increasing over time and is greater than one when the military spending multiplier is so. The weight for the non-military spending multiplier is the response of non-military spending to the news shock, which is negative and persistent. In contrast, the multipliers are statistically insignificant in recessions, as are shown in Figures 9 and 10. In sum, we find little statistical evidence supporting the state-dependency of the multipliers. Although point estimates of the non-military spending multipliers are larger than those in the low-unemployment state, the standard errors are too large to give a reasonable statistical significance.

Figure 11 shows the relationship between the military spending and the non-military spending multipliers. By replacing the population moments in (39)-(41) with the sample moments, we obtain the following equations:

$$\text{Linear:} \quad \widehat{\widehat{\theta}}_{h,y1} = \frac{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{t+h}^{\perp}}{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{1,t+h}^{\perp}} \widehat{m}_h - \frac{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{2,t+h}^{\perp}}{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{1,t+h}^{\perp}} M$$

$$\text{Low Unemployment:} \quad \widehat{\widehat{\theta}}_{h,y1} = \frac{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{t+h}^{\perp} 1(q_{t-1} \le \gamma^*)}{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{1,t+h}^{\perp} 1(q_{t-1} \le \gamma^*)} \widehat{m}_{A,h} - \frac{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{2,t+h}^{\perp} 1(q_{t-1} \le \gamma^*)}{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{1,t+h}^{\perp} 1(q_{t-1} \le \gamma^*)} M$$

$$\text{High Unemployment:} \quad \widehat{\widehat{\psi}}_{h,y1} = \frac{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{t+h}^{\perp} 1(q_{t-1} > \gamma^*)}{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{1,t+h}^{\perp} 1(q_{t-1} > \gamma^*)} \widehat{m}_{B,h} - \frac{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{2,t+h}^{\perp} 1(q_{t-1} > \gamma^*)}{\sum_{t=1}^{T} z_t^{\perp} \widetilde{x}_{1,t+h}^{\perp} 1(q_{t-1} > \gamma^*)} M,$$

where $\widehat{\widehat{\theta}}_{h,y1}$ is the non-military spending multiplier estimator and M is a given military spending multiplier. The first row shows the graphs of the 2 year integral case and the second row shows those of the 4 year integral case. The first column shows the linear case (no state-dependency) and the state-dependent cases all together. The second and third columns show the linear and state-dependent cases with 95% confidence bands, respectively. For both of 2 year and 4 year integrals, the non-military spending multiplier is greater than one when the military spending multiplier is greater than one.

It is worth noting that the linear case is very similar to the low unemployment case and not in the "middle" of the state-dependent cases. This is because the functions have similar slopes, but the intercepts of the linear and the high unemployment cases are quite different.

6 Conclusion

In this paper, we provide a formal theoretical framework for identification and estimation of dynamic causal effects in the LP-IV models. Our results are based on the structural vector moving average models. We show that the IV estimand is a weighted average of heterogeneous sectoral dynamic causal effects and provide conditions under which the sector-specific weights are identified. This result is extended to threshold models where the sample is split according to a continuous threshold variable. Since the correct specification of the threshold parameter is important for identification of weights, we propose a new test for state-dependency robust to heterogeneous sectoral dynamic causal effects. In contrast, the conventional test for state-dependency is no longer appropriate if heterogeneous sectoral dynamics are allowed. Finally we apply our method to estimate the non-military spending multiplier in the U.S. using the data after WWII. We first decompose the aggregate multiplier into a weighted sum of the military and non-military spending multipliers. The estimated aggregate spending multiplier is less than one, but our decomposition result shows that the non-military spending multiplier can be larger than one if the military spending multiplier is at least as large as one. However, we find no strong evidence on the state-dependency of the multiplier.

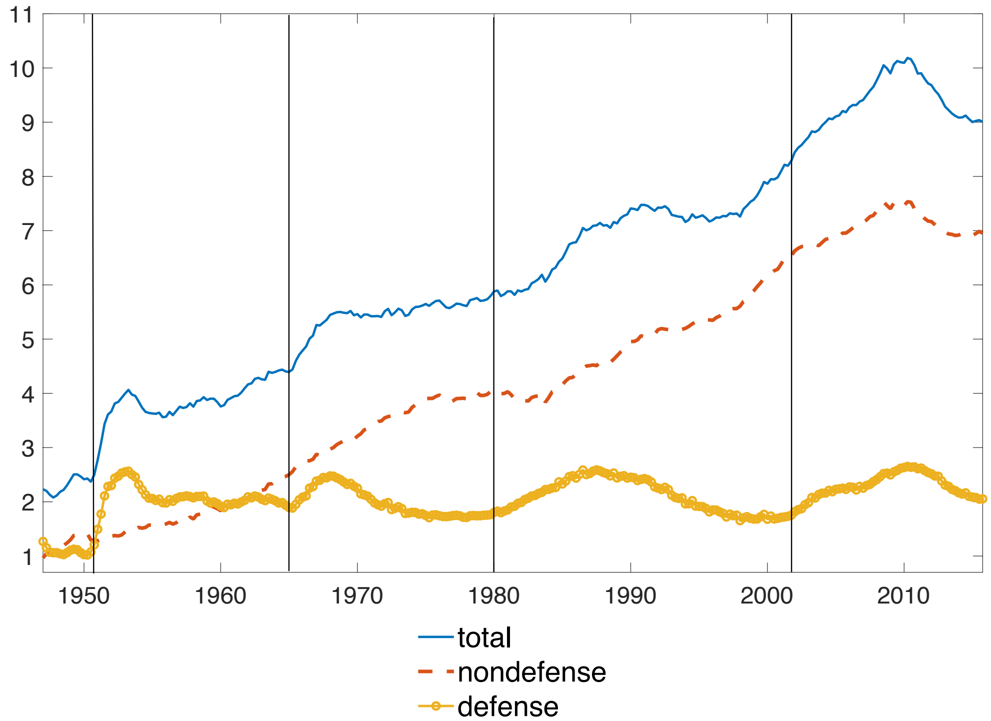

Figure 1: Real Government Spending Per Capita (in thousands of chained dollars)

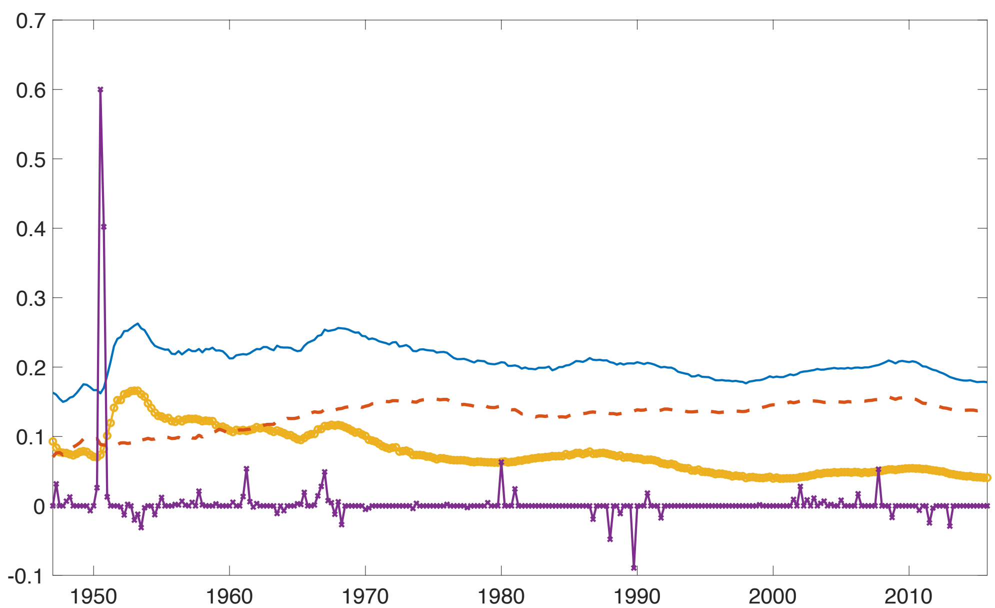

Figure 2: De-trended Real Government Spending and Military News Shock (Total spending: solid line, Defense spending: circle, Nondefense spending: dotted line, News shock: cross)

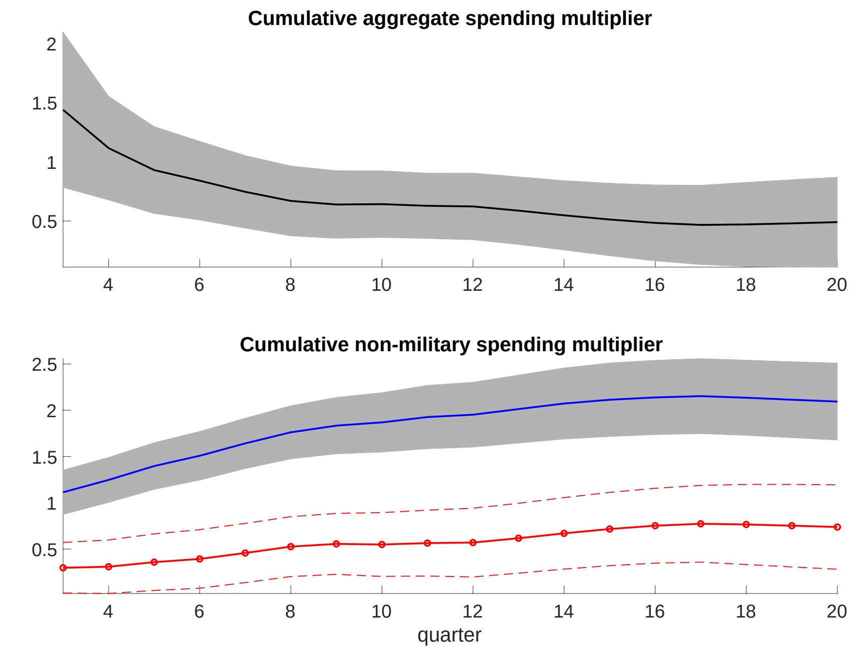

Figure 3: Cumulative Spending Multiplier with 95% Confidence Bands. Bottom panel: military spending multiplier = 1.2 (solid line), 0.6 (circle)

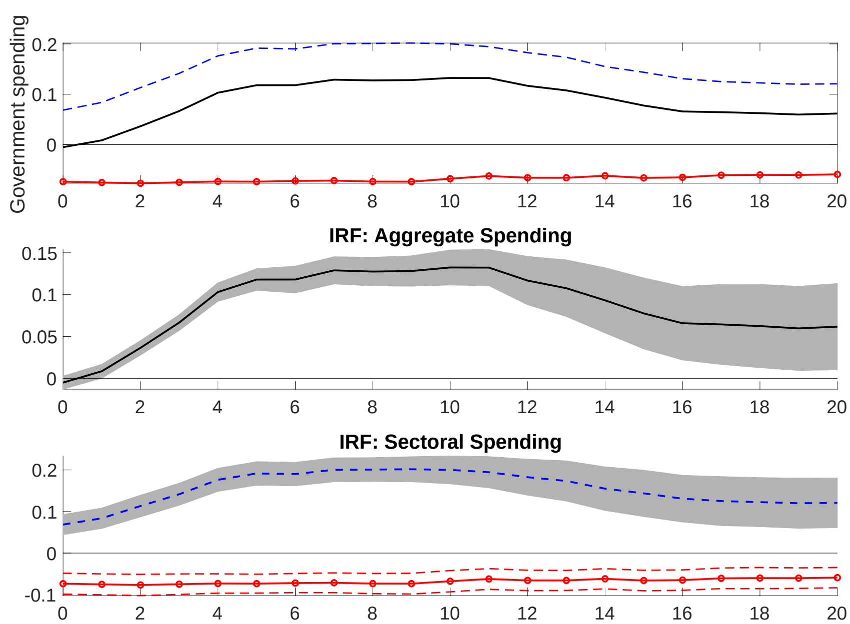

Figure 4: Government Spending Response to News Shock with 95% Confidence Bands. Bottom panel: military spending (solid line), non-military spending (circle)

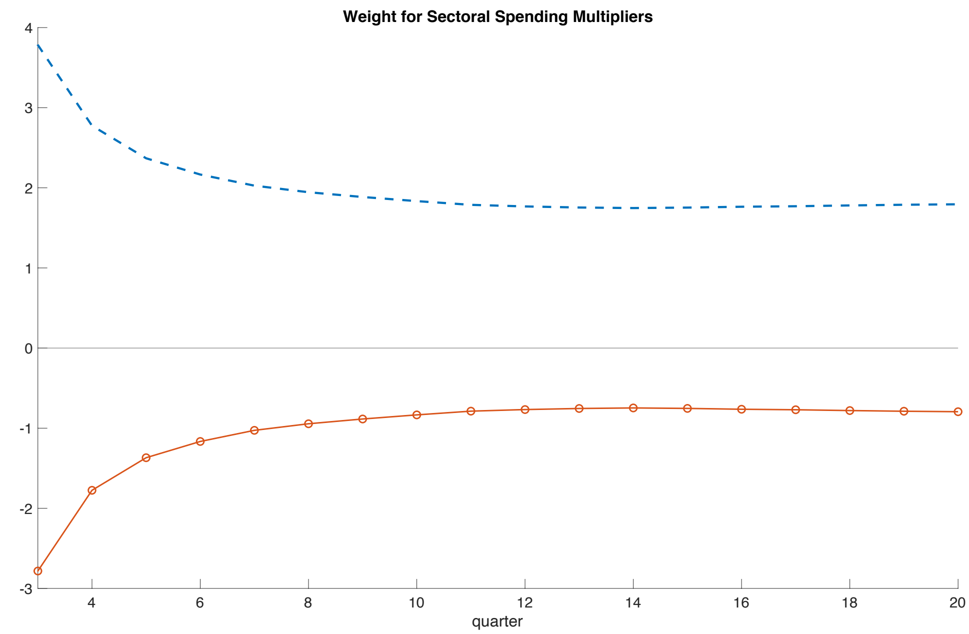

Figure 5: Relative Weights for Sectoral Multipliers: military spending multiplier (solid line), non-military spending multiplier (circle)

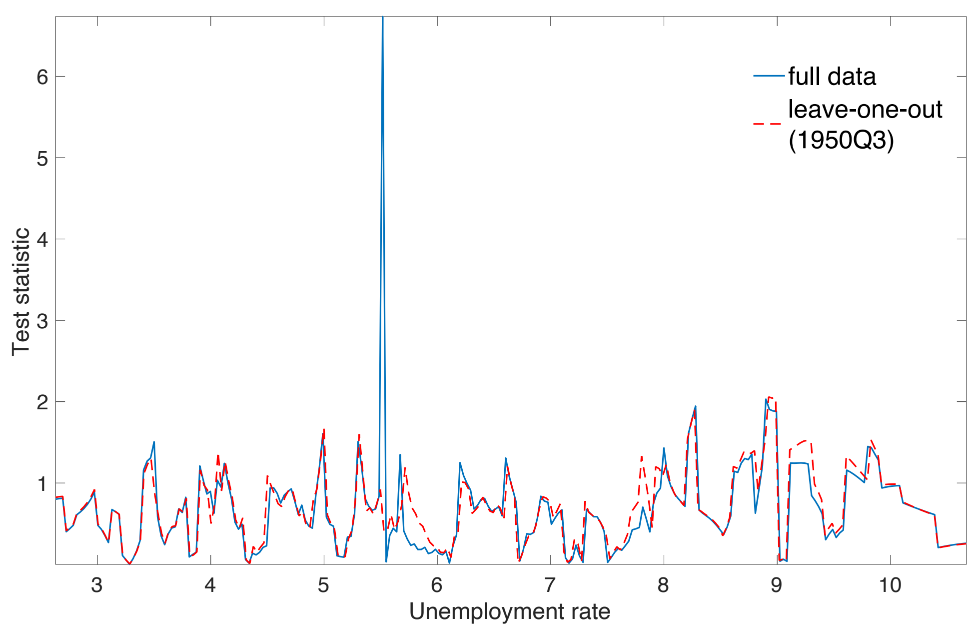

Figure 6: Test Statistic for Threshold Parameter using Post-WWII sample. Bootstrap 5% critical value=13.97

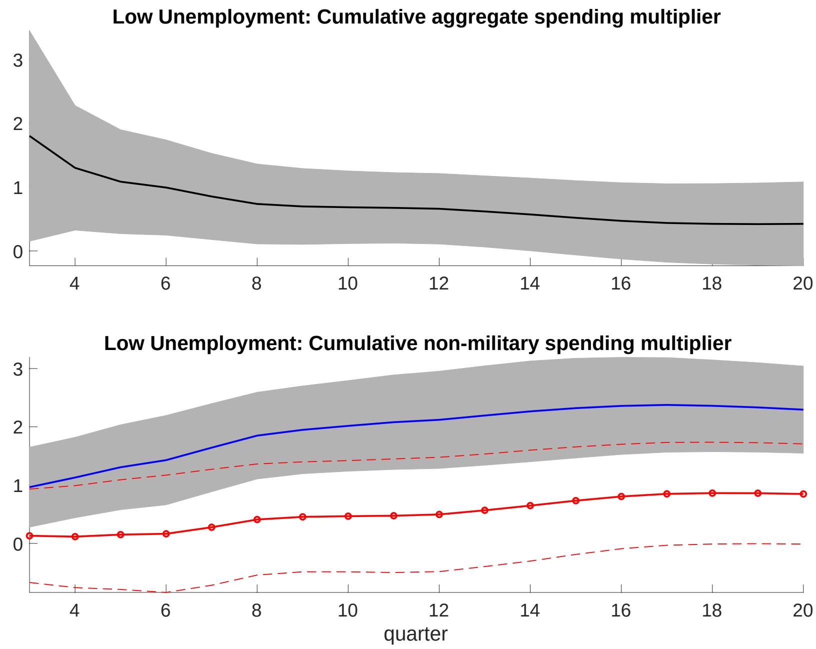

Figure 7: Cumulative Spending Multiplier with 95% Confidence Bands in Low-Unemployment State. Bottom panel: military spending multiplier = 1.2 (solid line), 0.6 (circle)

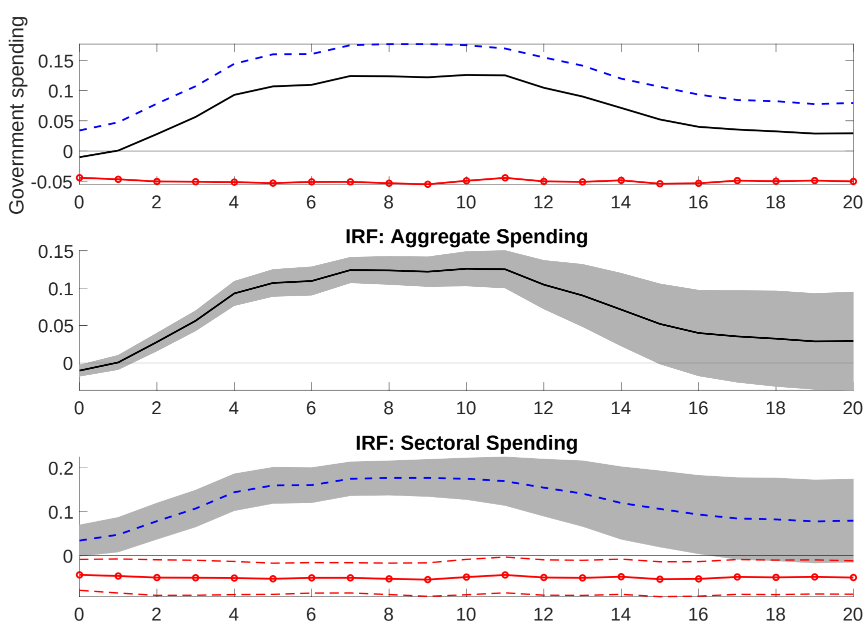

Figure 8: Government Spending Response to News Shock with 95% Confidence Bands in Low-Unemployment State. Bottom panel: military spending (solid line), non-military spending (circle)

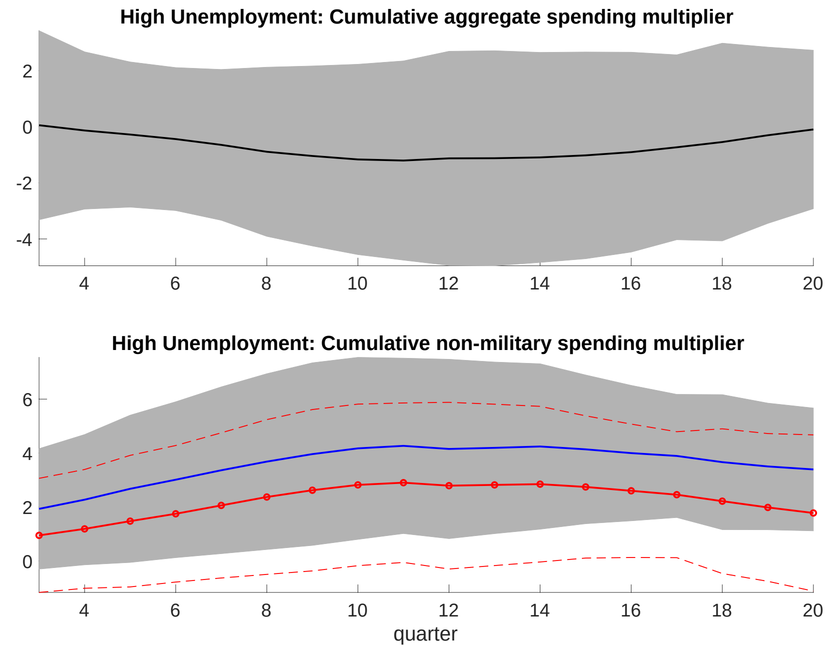

Figure 9: Cumulative Spending Multiplier with 95% Confidence Bands in High-Unemployment State. Bottom panel: military spending multiplier = 1.2 (solid line), 0.6 (circle)

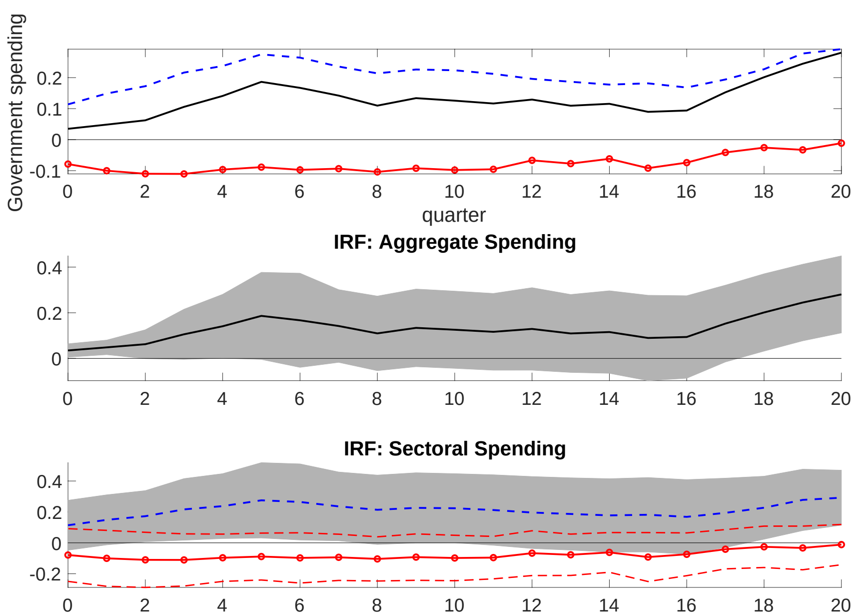

Figure 10: Government Spending Response to News Shock with 95% Confidence Bands in High-Unemployment State. Bottom panel: military spending (solid line), non-military spending (circle)

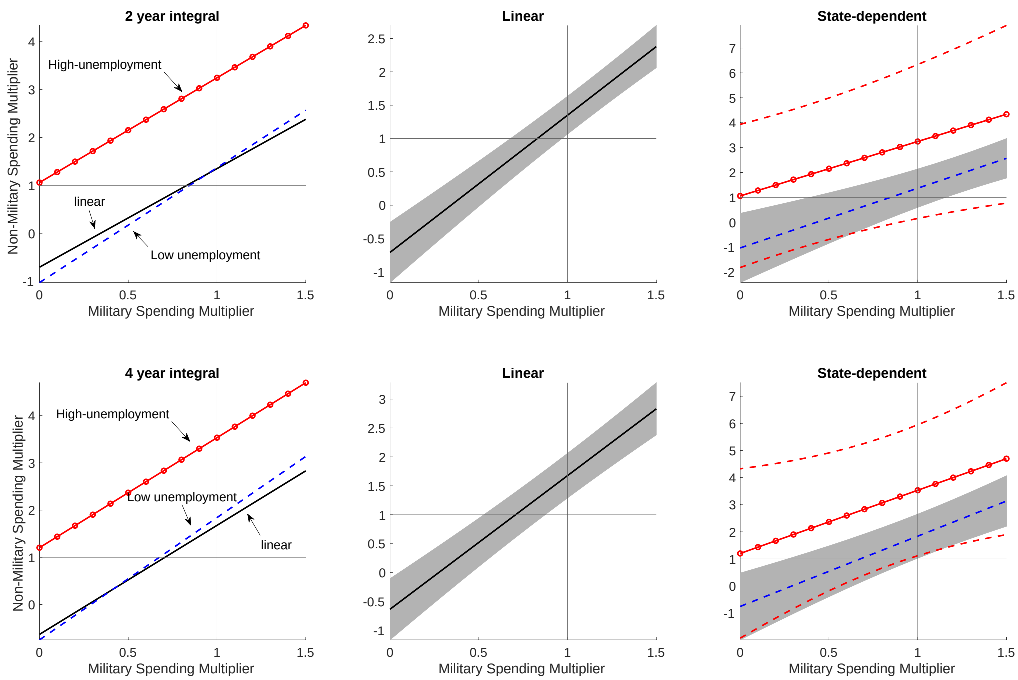

Figure 11: Non-Military Spending as a Function of Military Spending Multiplier

Appendix A: Proofs

A.1. Proof of Proposition 1

Rewriting (11) for y_t and $x_{s,t}$ yields

$$
\begin{aligned}
y_t &= (\varepsilon_{y,t} + \theta_{1,yy}\varepsilon_{y,t-1} + \theta_{2,yy}\varepsilon_{y,t-2} + \cdots) + \sum_{s=1}^{S} (\theta_{0,ys}\varepsilon_{s,t} + \theta_{1,ys}\varepsilon_{s,t-1} + \cdots) \\
&= \sum_{j=0}^{\infty} \left(\theta_{j,yy}\varepsilon_{y,t-j} + \sum_{s=1}^{S} \theta_{j,ys}\varepsilon_{s,t-j} \right) \qquad (42) \\
x_{s,t} &= (\theta_{0,sy}\varepsilon_{y,t} + \theta_{1,sy}\varepsilon_{y,t-1} + \theta_{2,sy}\varepsilon_{y,t-2} + \cdots) + \sum_{s'=1}^{S} \left(\theta_{0,ss'}\varepsilon_{s',t} + \theta_{1,ss'}\varepsilon_{s',t-1} + \cdots \right), \\
&= \sum_{j=0}^{\infty} \left(\theta_{j,sy}\varepsilon_{y,t-j} + \sum_{s'=1}^{S} \theta_{j,ss'}\varepsilon_{s',t-j} \right). \qquad (43)
\end{aligned}
$$

Replace the subscript t with $t+h$ in (42)-(43). Using $x_t = \sum_{s=1}^{S} x_{s,t}$ and Assumption 1, we obtain $E[z_t y_{t+h}] = \sum_{s=1}^{S} \alpha_s \theta_{h,ys}$ and $E[z_t x_{t+h}] = \sum_{s=1}^{S} \alpha_s \theta_{h,xs}$. Setting $h = 0$ for $E[z_t x_{t+h}]$ yields the estimand. Note that the denominator of the estimand is nonzero under Assumption 1(i). *Q.E.D.*

A.2. Proof of Proposition 2

Under the assumption, $\theta_{0,xs} = \theta_{0,ss} = 1$ by the unit effect normalization and the IV estimand simplifies to

$$\beta_h = \sum_{s=1}^{S} \frac{\alpha_s}{\sum_{s'=1}^{S} \alpha_{s'}} \theta_{h,ys}. \qquad (44)$$

Since

$$E[z_t x_{s,t}] = \sum_{s'=1}^{S} \alpha_{s'} \theta_{0,ss'} = \alpha_s \theta_{0,ss} = \alpha_s, \qquad (45)$$

the IV estimand can be written as

$$\beta_h = \sum_{s=1}^{S} \frac{E[z_t x_{s,t}]}{E[z_t x_t]} \theta_{h,ys}. \qquad (46)$$

Q.E.D.

A.3. Proof of Proposition 3

Similar to the proof of Proposition 1, we rewrite (14) for y_t and $x_{s,t}$ and replace t with $t+h$ to get

$$y_{t+h} = \sum_{j=0}^{\infty}(\theta_{j,yy}1(q_{t+h-1} \leq \gamma_0) + \psi_{j,yy}1(q_{t+h-1} > \gamma_0))\varepsilon_{y,t+h-j} \tag{47}$$
$$+ \sum_{j=0}^{\infty}\sum_{s=1}^{S}(\theta_{j,ys}1(q_{t+h-1} \leq \gamma_0) + \psi_{j,ys}1(q_{t+h-1} > \gamma_0))\varepsilon_{s,t+h-j},$$

$$x_{s,t+h} = \sum_{j=0}^{\infty}(\theta_{j,sy}1(q_{t+h-1} \leq \gamma_0) + \psi_{j,sy}1(q_{t+h-1} > \gamma_0)\varepsilon_{y,t+h-j} \tag{48}$$
$$+ \sum_{j=0}^{\infty}\sum_{s'=1}^{S}(\theta_{j,ss'}1(q_{t+h-1} \leq \gamma_0) + \psi_{j,ss'}1(q_{t+h-1} > \gamma_0))\varepsilon_{s',t+h-j}.$$

Using Assumption 2(ii)-(iii) and the law of iterated expectations, for all integer $j \leq h$,

$$E[z_t1(q_{t-1} \leq \gamma)1(q_{t+h-1} \leq \gamma_0)\varepsilon_{t+j}] = E[1(q_{t+h-1} \leq \gamma_0)E[1(q_{t-1} \leq \gamma)E[z_t\varepsilon_{t+j}|q_{t-1}]|q_{t+h-1}]] = 0,$$
$$E[z_t1(q_{t-1} \leq \gamma)1(q_{t+h-1} > \gamma_0)\varepsilon_{t+j}] = E[1(q_{t+h-1} > \gamma_0)E[1(q_{t-1} \leq \gamma)E[z_t\varepsilon_{t+j}|q_{t-1}]|q_{t+h-1}]] = 0,$$

and

$$E[z_t1(q_{t-1} \leq \gamma)1(q_{t-1} \leq \gamma_0)\varepsilon_{y,t}] = E[1(q_{t-1} \leq \gamma)1(q_{t-1} \leq \gamma_0)E[z_t\varepsilon_{y,t}|q_{t-1}]] = 0,$$
$$E[z_t1(q_{t-1} \leq \gamma)1(q_{t-1} > \gamma_0)\varepsilon_{y,t}] = E[1(q_{t-1} \leq \gamma)1(q_{t-1} > \gamma_0)E[z_t\varepsilon_{y,t}|q_{t-1}]] = 0.$$

In addition, we can write

$$E[z_t1(q_{t-1} \leq \gamma)1(q_{t-1} \leq \gamma_0)\varepsilon_{s,t}] = \begin{cases} \alpha_{A,s}(\gamma) & \text{if } \gamma \leq \gamma_0, \\ \alpha_{A,s}(\gamma_0) & \text{if } \gamma > \gamma_0, \end{cases}$$
$$E[z_t1(q_{t-1} \leq \gamma)1(q_{t-1} > \gamma_0)\varepsilon_{s,t}] = \begin{cases} 0 & \text{if } \gamma \leq \gamma_0, \\ \alpha_s(\gamma_0, \gamma) & \text{if } \gamma > \gamma_0. \end{cases}$$

Combining the above results, we obtain

$$E[z_t1(q_{t-1} \leq \gamma)y_{t+h}] = \begin{cases} \sum_{s=1}^{S}\alpha_{A,s}(\gamma)\theta_{h,ys} & \text{if } \gamma \leq \gamma_0, \\ \sum_{s=1}^{S}(\alpha_{A,s}(\gamma_0)\theta_{h,ys} + \alpha_s(\gamma_0, \gamma)\psi_{h,ys}) & \text{if } \gamma > \gamma_0, \end{cases}$$
$$E[z_t1(q_{t-1} \leq \gamma)x_{s,t+h}] = \begin{cases} \sum_{s'=1}^{S}\alpha_{A,s'}(\gamma)\theta_{h,ss'} & \text{if } \gamma \leq \gamma_0, \\ \sum_{s'=1}^{S}(\alpha_{A,s'}(\gamma_0)\theta_{h,ss'} + \alpha_{s'}(\gamma_0, \gamma)\psi_{h,ss'}) & \text{if } \gamma > \gamma_0 \end{cases}$$

for $\beta_{A,h}(\gamma)$. For $\beta_{B,h}(\gamma)$, we use

$$E[z_t 1(q_{t-1} > \gamma) 1(q_{t-1} \leq \gamma_0)\varepsilon_{s,t}] = \begin{cases} \alpha_s(\gamma, \gamma_0) & \text{if } \gamma \leq \gamma_0, \\ 0 & \text{if } \gamma > \gamma_0, \end{cases}$$

$$E[z_t 1(q_{t-1} > \gamma) 1(q_{t-1} > \gamma_0)\varepsilon_{s,t}] = \begin{cases} \alpha_{B,s}(\gamma_0) & \text{if } \gamma \leq \gamma_0, \\ \alpha_{B,s}(\gamma) & \text{if } \gamma > \gamma_0. \end{cases}$$

which lead to

$$E[z_t 1(q_{t-1} > \gamma) y_{t+h}] = \begin{cases} \sum_{s=1}^{S} (\alpha_{B,s}(\gamma_0)\psi_{h,ys} + \alpha_s(\gamma, \gamma_0)\theta_{h,ys}) & \text{if } \gamma \leq \gamma_0, \\ \sum_{s=1}^{S} \alpha_{B,s}(\gamma)\psi_{h,ys} & \text{if } \gamma > \gamma_0, \end{cases}$$

$$E[z_t 1(q_{t-1} > \gamma) x_{s,t+h}] = \begin{cases} \sum_{s'=1}^{S} (\alpha_{B,s'}(\gamma_0)\psi_{h,ss'} + \alpha_{s'}(\gamma, \gamma_0)\theta_{h,ss'}) & \text{if } \gamma \leq \gamma_0, \\ \sum_{s'=1}^{S} \alpha_{B,s'}(\gamma)\psi_{h,ss'} & \text{if } \gamma > \gamma_0. \end{cases}$$

Now by setting $h = 0$ for $x_{s,t+h}$ and using $x_t = \sum_{s=1}^{S} x_{s,t}$, we have $\beta_{A,h}(\gamma)$ and $\beta_{B,h}(\gamma)$. *Q.E.D.*

A.4. Proof of Proposition 4

Under the assumption, $\theta_{0,xs} = \theta_{0,ss} = 1$ and $\psi_{0,xs} = \psi_{0,ss} = 1$ by the unit effect normalization. Letting $\gamma = \gamma_0$, the IV estimands simplify to

$$\beta_{A,h}(\gamma_0) = \sum_{s=1}^{S} \left(\frac{\alpha_{A,s}(\gamma_0)}{\sum_{s=1}^{S} \alpha_{A,s}(\gamma_0)} \right) \theta_{h,ys}, \tag{49}$$

$$\beta_{B,h}(\gamma_0) = \sum_{s=1}^{S} \left(\frac{\alpha_{B,s}(\gamma_0)}{\sum_{s=1}^{S} \alpha_{B,s}(\gamma_0)} \right) \psi_{h,ys}. \tag{50}$$

Since

$$E[z_t 1(q_{t-1} \leq \gamma_0) x_{s,t}] = \sum_{s'=1}^{S} \alpha_{A,s'}(\gamma_0)\theta_{0,ss'} = \alpha_{A,s}(\gamma_0)\theta_{0,ss} = \alpha_{A,s}(\gamma_0), \tag{51}$$

$$E[z_t 1(q_{t-1} > \gamma_0) x_{s,t}] = \sum_{s'=1}^{S} \alpha_{B,s'}(\gamma_0)\psi_{0,ss'} = \alpha_{B,s}(\gamma_0)\psi_{0,ss} = \alpha_{B,s}(\gamma_0), \tag{52}$$

the IV estimands can be written as desired. *Q.E.D.*

A.5. Proof of Theorem 1

Recall that the test statistic is given as in (33). Note that the major difference between the one in Proposition 2 and the test statistic is the presence of estimator of ν. To address the estimation error of ν, we refer to the result in Chernozhukov et al. (2014b). Particularly, Provided that Condition H4 in page 1795 of the paper is satisfied, then the estimation error does not affect the rate and can be treated as if it were known a priori. The type of estimation error in this paper is verified in pages 1813-1814 in Chernozhukov et al. (2014b). Provided that $K_b^{\pm}(\cdot)$ satisfies Assumption 4(ii), the desired result comes from the straightforward application of Proposition 2. *Q.E.D.*

Appendix B

Our distribution theory entails approximating suprema of local empirical processes with dependent data by suprema of Gaussian processes (strong approximation). This section discusses asymptotic theory on strong approximation in the time series context in details. Before our discussion, we begin with defining two concepts.

Definition 1. *A β-mixing coefficient is defined as*

$$\beta(k) := \sup_{t \geq 1} E \sup_{A \in \sigma_{t+k}^{\infty}} |Pr(A|\sigma_{-\infty}^{t}) - Pr(A)|$$

Definition 2. *Let $\mathcal{F}$ be a class of measurable functions on a measurable space $(S, \mathcal{S})$, to which a measurable envelope $\mathbf{F}$ is attached. Define $\|f\|_{Q,p} := (Q|f|^p)^{1/p} < \infty$. $\mathcal{F}$ is VC class with envelope $\mathbf{F}$ if there are constants $A, v > 0$ such that*

$$\sup_{Q} N(\varepsilon \|\mathbf{F}\|_{Q,2}, \mathcal{F}, e_Q) \leq (A/\varepsilon)^v$$

for all $0 < \varepsilon \leq 1$, where $e_Q(f, g) = \|f - g\|_{Q,2}$, $f, g \in L_2(Q)$, the supremum is taken over all finitely discrete probability measures on $(S, \mathcal{S})$.

B.1 Strong approximation for empirical processes of dependent data

Given the test statistic defined as (33), consider the following maxima of the empirical process of dependent data $\{(q_{t-1}, g_t)\}_{t=1}^{T}$:

$$W_T := \sup_{f \in \mathcal{F}_T} \mathbb{G}_T f, \tag{53}$$

where

$$\mathbb{G}_T f = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(f_q(q_{t-1}, g_t) - E f_q(q_{t-1}, g_t) \right) \tag{54}$$

and

$$\mathcal{F}_T = \left\{ f_q(x, y) : \left(K_b^{+}(q - x) - K_b^{-}(q - x) \right) y / \sqrt{b} \text{ where } q \in \Gamma \right\}. \tag{55}$$

Defining a two by one vector $w_t := (q_{t-1}, g_t)^{\top}$, $f_q(w_t) \in \mathcal{F}_T$ is indexed by q where Γ is a Borel subset of $\mathbb{R}$.[3] A sequence of random pair $\{w_t\}_{t=1}^{T}$ is assumed to be strictly stationary β-mixing for empirical processes $\mathbb{G}_T f$. Since $\mathcal{F}_T$ needs not belong to a Donsker class (while it belongs to a Vapnik-Červonenkis (VC) subgraph class), the usual weak convergence for W_T does not hold and therefore, strong approximation for W_T is required.[4]

[3] Without loss of generality, we can let $\Gamma = (b, 1 - b)$, which can be done by defining the conditioning variable by $\arctan(q_t)$.

[4] Note that the set of all translates of kernel function $K(\cdot)$ of bounded variation is VC and for two VC-subgraphs $\mathcal{F}$ and $\mathcal{G}$ with $f \in \mathcal{F}$ and $g \in \mathcal{G}$, a variety of operations including $\mathcal{F} \wedge \mathcal{G}$ and $\mathcal{F} \vee \mathcal{G}$, $\mathcal{F} + g$ and $\mathcal{F}g$ are VC-subgraphs. See e.g. Lemma (2.6.11) and Lemma (2.6.18) in van der Vaart and Wellner (2011). Some general sufficient conditions to ensure that $\mathcal{F}_T$ is VC are discussed in Giné and Guillou (2002).

Under this setting, by employing an entropy method, we show the asymptotic behavior of the test statistic by validating a Gaussian approximation of W_T by a sequence of random variables $\mathcal{W}_T$ whose distribution is equal to that of $\sup_{f\in\mathcal{F}_T} \mathrm{G}_T f$ where G_T is a centered Gaussian process specified below in Proposition 2 in a sense that

$$|W_T - \mathcal{W}_T| = o_p(\rho_T) \tag{56}$$

for some $\rho_T \to 0$ as $T \to \infty$.

While Chernozhukov, Chertverikov and Kato (2014) studied the above strong invariance principle for i.i.d data, their results cannot be directly applied to our test statistic since the data $\{w_t\}_{t=1}^T$ are dependent by nature. In order to extend Theorem 2.1 and Corollary 2.2 in Chernozhukov, Chertverikov and Kato (2014) by accommodating dependent data, we employ the Bernstein's big-block-small-block decomposition (e.g. Masry, 2005), coupling (e.g. Chen and Shen, 1998) and Martingale difference sequence (MDS) approximation (Wu and Mielniczuk, 2002).

In particular, following Bernstein's big-block-small-block decomposition to the empirical processes, we consider two sequences m and $r(< m)$ such that $m + r \le T/2$, $m, r \to \infty$ as $T \to \infty$, $m = o(T)$, and $r = o\left(m^{1/4}/\log^{1/2} T\right)$. Then we can construct two types of index sets: one type is a big block and the other is a small block such that for k_T satisfying $(k_T - 1)(m + r) \le T \le k_T(m + r)$, for $i = 1, ..., k_{T-1}$,

$$\begin{aligned}
I_i &= \{(i-1)(m+r)+1, ..., (i-1)(m+r)+m\}, \\
J_i &= \{im + (i-1)r + 1, i(m+r)\},
\end{aligned}$$

and the remainder for each type:

$$\begin{aligned}
I_{k_T} &= \{(k_T-1)(m+r)+1, ..., (k_T-1)(m+r)+m\} \cap \{1, ..., T\}, \\
J_{k_T} &= \{k_T m + (k_T-1)r + 1, k_T(m+r)\} \cap \{1, ..., T\}.
\end{aligned}$$

Based on these index sets, we can construct a random sequence $\{\tilde{w}_t\}_{t=1}^T$ independent of $\{w_t\}_{t=1}^T$ where $\{\tilde{w}_t\}$ is the sequence in the independent blocks such that $\{\tilde{w}_t : t \in I_i\}$ is independent of $\{\tilde{w}_t : t \in I_j\}$ and $\{\tilde{w}_t : t \in J_i\}$ is independent of $\{\tilde{w}_t : t \in J_j\}$ for $i \ne j$ satisfying that for any $M > 0$,

$$\begin{aligned}
&\left| P\left\{\sup_f \left|\frac{1}{\sqrt{T}}\sum_{i=1}^{k_T}\sum_{t\in I_i} f(w_t) - Ef(w_t)\right| \ge M\right\} - P\left\{\sup_f \left|\frac{1}{\sqrt{T}}\sum_{i=1}^{k_T}\sum_{t\in I_i} f(\tilde{w}_t) - Ef(\tilde{w}_t)\right| \ge M\right\}\right| \\
&= O(k_T \beta(r)),
\end{aligned}$$

$$\left| P\left\{ \sup_f \left| \frac{1}{\sqrt{T}} \sum_{i=1}^{k_T} \sum_{t\in J_i} f\left(w_t\right) - Ef\left(w_t\right) \right| \geq M \right\} - P\left\{ \sup_f \left| \frac{1}{\sqrt{T}} \sum_{i}^{k_T} \sum_{t\in J_i} f\left(\tilde{w}_t\right) - Ef\left(\tilde{w}_t\right) \right| \geq M \right\} \right|$$
$$= O\left(k_T \beta\left(m\right)\right),$$

due to the coupling as in e.g. Lemma 1 and its proof of Chen and Shen (1998). Then, the above approximation errors are asymptotically negligible as long as

$$k_T \beta\left(r\right) = o\left(T\beta\left(m^{1/4}/\log^{1/2} T\right)/m\right) \to 0. \tag{57}$$

where $m = T^c$ with $0 < c < 1$. And $k_T \beta\left(m\right) = o\left(1\right)$ since $r = o\left(m\right)$. It suffices to set a polynomial decay rate for $\beta\left(\cdot\right)$, e.g. $\beta(k) = o(k^{-\eta})$ for some positive constant $\eta > 4/c - 4$.

Consider the following collections of composite functions:

$$\begin{aligned} \mathcal{U}_1 =& \{\ell_{1,q}\left(\tilde{w}_t : t \in I_i\right) = \sum_{t\in I_i} f_q(\tilde{w}_t) : q \in \Gamma, f_q \in \mathcal{F}_T\} \\ \mathcal{U}_2 =& \{\ell_{2,q}\left(\tilde{w}_t : t \in J_i\right) = \sum_{t\in J_i} f_q(\tilde{w}_t) : q \in \Gamma, f_q \in \mathcal{F}_T\} \end{aligned}$$

Under the blocking and coupling scheme, $\mathbb{G}_T f = \mathbb{G}_{1,k_T}\ell_1 + \mathbb{G}_{2,k_T}\ell_2$ where $\ell_1 \in \mathcal{U}_1$ and $\ell_2 \in \mathcal{U}_2$ and $\mathbb{G}_{1,k_T}\ell_1$ is an empirical process indexed by $\mathcal{U}_1$ such that

$$\mathbb{G}_{1,k_T}\ell_1 = \frac{1}{\sqrt{T}} \sum_{k=1}^{k_T} \left(\ell_{1,q}(\tilde{w}_t : t \in I_k) - E\ell_{1,q}(\tilde{w}_t : t \in I_k)\right) \tag{58}$$

and $\mathbb{G}_{2,T}\ell_2$ is defined analogously.

Before we proceed with our detailed discussion on the strong approximation, we provide regularity conditions.

Assumption 4.

(i) $E|g_t|^p < \infty$ *for some* $p \geq 4$ *and* $\sup_{q\in\Gamma} E[g_t^4 | q_{t-1} = q] < \infty$.

(ii) The kernel function $K(\cdot)$ *is a bounded and continuous on* $\mathbb{R}$ *and such that the class of functions* $K(\cdot)$ *belongs to is a VC class with a bounded envelop.*

(iii) The distribution of q_t *has a bounded Lebesgue density* $p_q(\cdot)$ *on* $\mathbb{R}$.

(iv) $C_\Gamma := \sup_{T\geq 1} \sup_{q\in\Gamma} |\nu^{-1/2}(q)| < \infty$. *Moreover, for every fixed* $T \geq 1$ *and for every* $q_k \in \Gamma$ *with* $q_k \to q \in \Gamma$ *pointwise,* $\nu(q_k) \to \nu(q)$.

(v) For the bandwidth b*, the big block size* m*, and the small block size* r*, as* $T \to \infty$*, (a)* $b \to 0$ *(b)* $m = T^c$ *for* $0 < c < 1$*; (c)* $r = m^{1/4}/\log^{1/2} T$.

(vi) $\{w_t\}$ *is a strictly stationary* β*-mixing sequence with its* β*-mixing coefficient satisfying* $\beta(k) = o(k^{-\eta})$ *for some constant* $\eta > 4/c - 4$ *for* c *in (v).*

(vii) $\{w_t\}$ *is a linear process with the absolute summability and finite second moment such that* $\tilde{w}_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}$ *where* $\sum_i^{\infty} |a_i| < \infty$ $\{\epsilon_i\}$ *are i.i.d random variables with zero mean and a finite variance and the density of* ϵ_1 *is Lipschitz continuous and bounded.*

Assumptions 4(i)-(v) are slightly modified from those for Proposition 3.2. in Chernozhukov et al. (2014). Assumption 4(iii) is typical and Assumption 4(iv) is required to avoid measurability complications. Assumption 4(v) is the decaying rate for the bandwidth and the increasing rate for the size of the big block, and the size of the small block. Assumption 4(vi) allows for stationary data and it is required to relate the original empirical process with dependent data to another empirical process built on an independent block sequence along the line of blocking and coupling scheme in Arcone and Yu (1994) and Chen and Shen (1998). Assumption 4(vii) is required for the MDS approximation in relation to the big block component as in Wu and Mielniczuk (2002).

The following lemma shows the uniform asymptotic negligibility of the small block components ($\mathbb{G}_{2,k_T}\ell_2$) compared to the big block ones ($\mathbb{G}_{1,k_T}\ell_1$) via a maximal inequality for the empirical processes so that we can focus on the big block components.

Lemma 1. *Under Assumption 4,*

$$\sup_{\ell_2 \in \mathcal{U}_2} |\mathbb{G}_{2,k_T}\ell_2| = o_p(\log^{-1} T)$$

Proof. The size of envelop for the class $\mathcal{U}_2$ is given as $\left\| \sum_{t \in J_k} F(\tilde{w}_t) \right\|_{P,2} \leq r \|F\|_{P,2}$ and the uniform entropy integral has

$$\sup_Q \int_0^1 \sqrt{1 + \log N \left(\varepsilon \left\| \sum_{t \in J_k} F(\tilde{w}_t) \right\|_{Q,2}, \mathcal{U}_2, L_2(Q) \right)} d\varepsilon = O(r).$$

Apply Theorem 2.14.1 in van der Vaart and Wellner (2011) or Corollary 5.1 in Chernozhukov et al. (2014) to the class $\mathcal{U}_2$,

$$\sup_{\ell_2 \in \mathcal{U}_2} |\mathbb{G}_{2,k_T}\ell_2| = \sup_{f \in \mathcal{F}} \left| \frac{\sqrt{k_T}}{\sqrt{T}} \frac{1}{\sqrt{k_T}} \sum_{k=1}^{k_T} \sum_{t \in J_k} f(\tilde{w}_t) - Ef(\tilde{w}_t) \right| = O_p\left(\frac{r^2}{\sqrt{m+r}} \right) = o_p\left(\log^{-1} T\right)$$

due to Assumption 4(v)(c). *Q.E.D.*

For the big block component, note that while the big blocks in the sequence, $\{\tilde{w}_t\}$ are independent, elements within each block are dependent. To address the dependence within each big block, we employ the martingale approximation following Lemma 3 in Wu and Mielniczuk (2002). In particular, the following Lemma shows that the big block component can be approximated by an MDS under Assumption 4.

Lemma 2. *Under Assumption 4,*

$$\sup_{\ell_1 \in \mathcal{U}_1} \left| \mathbb{G}_{1,k_T} \ell_1(\tilde{w}_t) - \mathbb{G}_{1,k_T} \ell_1(\tilde{\xi}_t) \right| \tag{59}$$

$$= \sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{T}} \sum_{k=1}^{k_T} \sum_{t \in I_k} \left[f\left(\tilde{w}_t\right) - Ef\left(\tilde{w}_t\right) \right] - \frac{1}{\sqrt{T}} \sum_{k=1}^{k_T} \sum_{t \in I_k} [f(\tilde{\xi}_t) - Ef(\tilde{\xi}_t)] \right| = o_p\left(\sqrt{b}\right), \tag{60}$$

where $\{\tilde{\xi}_t\}$ is an MDS and the approximation error is shown to be of the order of $O\left(\sqrt{b}\right)$, where b is the bandwidth in the kernel.

Proof. Applying Lemma 3 in Wu and Mielniczuk (2002) yields the desired result. *Q.E.D.*

Since each small block is asymptotically negligible due to Lemma 1 and each big block in the sequence $\{\tilde{w}_t\}$ is approximated by an MDS as in (60) due to Lemma 2, we can apply Corollary 2.2 in Chernozhukov et al. (2014) to the independent blocks.

Theorem 2. *Under Assumption 4, for W_T defined in (53), there exists $\mathcal{W}_T$ whose distribution is equivalent to that of $\sup_{f \in \mathcal{F}_T} G_T f$, a sequence of suprema of independent centered Gaussian processes G_T indexed by Γ where $q \in \Gamma$ with covariance function such that for $f_q, f_{q'} \in \mathcal{F}_T$,*

$$E[\mathrm{G}_T f_q(\tilde{\xi}_t) G_T f_{q'}(\tilde{\xi}_t)] = Cov\left(\frac{1}{\sqrt{mk_T}} \sum_{j=1}^{k_T} \sum_{t \in I_j} f_q\left(\tilde{\xi}_t\right), \frac{1}{\sqrt{mk_T}} \sum_{j=1}^{k_T} \sum_{t \in I_j} f_{q'}\left(\tilde{\xi}_t\right) \right), \tag{61}$$

where $\tilde{\xi}_t$ is a sequence of MDS. Then,

$$|W_T - \mathcal{W}_T| = O_p(A_T + B_T + C_T) \tag{62}$$

with

$$\begin{aligned} A_T &= m^{1/2-1/q} b^{-1/2} T^{-1/2+1/q} \log^{3/2} T \\ B_T &= m^{1/4} \left(bT\right)^{-1/4} \log^{5/4} T \\ C_T &= \left(bT\right)^{-1/6} \log T \end{aligned}$$

where m is the size of each big block, b is the bandwidth.

Proof. Due to Lemma 2, we first apply Corollary 2.2 in Chernozhukov et al. (2014) to the big block component with $\{\tilde{\xi}_t\}$. In particular, we set the parameters in Corollary 2.2 in Chernozhukov et al. (2014) as follows: $b = mb^{-1/2}$, $\sigma = O\left(1\right)$, $K_n = \log T$ and $\mathcal{F} = \left\{ \sum_{t \in I_k} \frac{f\left(b^{-1}(q - q_{t-1})\right)}{\sqrt{b}} : q \in \Gamma \right\}$. As $k_T\left(m + r\right) \sim T$ and $r = o\left(m\right)$, we have $k_T m/n \to 1$. Then, there exists a sequence of suprema of independent centered Gaussian processes G_{k_T} indexed by Γ where $q \in \Gamma$ with covariance function such that for $\ell_{1,q}, \ell_{1,q'} \in \mathcal{U}_1$,

$$E[\mathrm{G}_{k_T}\ell_{1,q}\mathrm{G}_{k_T}\ell_{1,q'}] = \mathrm{Cov}\left(\frac{1}{\sqrt{k_T}}\sum_{j=1}^{k_T}\sum_{t\in I_j} f_q\left(\tilde{\xi}_t\right), \frac{1}{\sqrt{k_T}}\sum_{j=1}^{k_T}\sum_{t\in I_j} f_{q'}\left(\tilde{\xi}_t\right)\right). \tag{63}$$

Then, applying Corollary 2.2 in Chernozhukov et al. (2014) yields

$$\begin{aligned}
&\frac{1}{\sqrt{m}}\left(\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{k_T}}\sum_{j=1}^{k_T}\sum_{t\in I_j} f\left(\tilde{\xi}_t\right) - \sup_{f\in\mathcal{F}}\mathrm{G}_{k_T}\ell_{1,q}\right) \\
&\quad= \frac{1}{\sqrt{m}}O_p\left(mb^{-1/2}k_T^{-1/2+1/q}\log^{3/2}T + m^{1/2}b^{-1/4}k_T^{-1/4}\log^{5/4}T + m^{1/3}\left(bk_T\right)^{-1/6}\log T\right) \\
&\quad= \frac{1}{\sqrt{m}}O_p\left(m^{3/2-1/q}b^{-1/2}T^{-1/2+1/q}\log^{3/2}T + m^{3/4}\left(bT\right)^{-1/4}\log^{5/4}T + m^{1/2}\left(bT\right)^{-1/6}\log T\right).
\end{aligned}$$

Also, due to Lemma 1,

$$\frac{1}{\sqrt{m}}\sup_{f\in\mathcal{F}_T}\mathrm{G}_{k_T}f = \sup_{f\in\mathcal{F}_T}\mathrm{G}_T f + o_p\left(\log^{-1}T\right).$$

Combining the above result yields the desired result. *Q.E.D.*

References

[1] Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. Journal of Econometrics, 113(2), 231-263.

[2] Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American Statistical Association, 90(430), 431-442.

[3] Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91(434), 444-455.

[4] Angrist, J. D., Jord, ., & Kuersteiner, G. M. (2018). Semiparametric estimates of monetary policy effects: string theory revisited. Journal of Business & Economic Statistics, 36(3), 371-387.

[5] Antoch, J., Gregoire, G., & Hukov, M. (2007). Tests for continuity of regression functions. Journal of Statistical Planning and Inference, 137(3), 753-777.

[6] Arcones, M. A., & Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. Journal of Theoretical Probability, 7(1), 47-71.

[7] Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. American Economic Journal: Economic Policy, 4(2), 1-27.

[8] Barro, R. J., & Redlick, C. J. (2011). Macroeconomic effects from government purchases and taxes. The Quarterly Journal of Economics, 126(1), 51-102.

[9] Blanchard, O., & 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.

[10] Caner, M., & Hansen, B. E. (2004). Instrumental variable estimation of a threshold model. Econometric Theory, 813-843.

[11] Chen, X., & Shen, X. (1998). Sieve extremum estimates for weakly dependent data. Econometrica, 289-314.

[12] Chernozhukov, V., Chetverikov, D., & Kato, K. (2014). Gaussian approximation of suprema of empirical processes. The Annals of Statistics, 42(4), 1564-1597.

[13] Chernozhukov, V., Chetverikov, D., & Kato, K. (2014b). Anti-concentration and honest, adaptive confidence bands. The Annals of Statistics, 42(5), 1787-1818.

[14] Gin, E., & Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincare (B) Probability and Statistics, 38(6), 907-921.

[15] Hansen, B. E. (2020). Econometrics. unpublished manuscript. University of Wisconsin-Madison.

[16] Heckman, J. J., Urzua, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. The Review of Economics and Statistics, 88(3), 389-432.

[17] Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. Econometrica, 73(3), 669-738.

[18] Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. Econometrica, 62(2), 467-475.

[19] Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. American Economic Review, 95(1), 161-182.

[20] Jord, ., Schularick, M., & Taylor, A. M. (2020). The effects of quasi-random monetary experiments. Journal of Monetary Economics, 112, 22-40.

[21] Jordà, Ò., & Taylor, A. M. (2016). The time for austerity: estimating the average treatment effect of fiscal policy. The Economic Journal, 126(590), 219-255.

[22] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. Stochastic Processes and their Applications, 115(1), 155-177.

[23] Nakamura, E., & Steinsson, J. (2014). Fiscal stimulus in a monetary union: Evidence from US regions. American Economic Review, 104(3), 753-92.

[24] Plagborg-Mller, M., & Wolf, C. K. (forthcoming). Local projections and VARs estimate the same impulse responses. Econometrica.

[25] Plagborg-Mller, M., & Wolf, C. K. (2020). Instrumental variable identification of dynamic variance decompositions. arXiv preprint arXiv:2011.01380.

[26] Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. The Quarterly Journal of Economics, 126(1), 1-50.

[27] Ramey, V. A., & Shapiro, M. D. (1998). Costly capital reallocation and the effects of government spending. Carnegie-Rochester conference series on public policy, Vol. 48, 145-194. North-Holland.

[28] Seo, M. H., & Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. Journal of Econometrics, 195(2), 169-186.

[29] Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. The Economic Journal, 128(610), 917-948.

[30] Ramey, V. A., & 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.

[31] van der Vaart, A., & Wellner, J. A. (2011). A local maximal inequality under uniform entropy. Electronic Journal of Statistics, 5(2011), 192.

[32] Wu, W. B., & Mielniczuk, J. (2002). Kernel density estimation for linear processes. The Annals of Statistics, 30(5), 1441-1459.