Online Supplement for "Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly"

Michal Kolesár
Princeton University

Mikkel Plagborg-Møller
Princeton University

July 15, 2025

Appendix A Further results

A.1 Integrals of the weight function

The following lemma provides an identification result for integrals of the causal weight function ω_X defined in Section 3.1.

Lemma A.1. *Let ω_X be given by (9). Assume that $E[X_t^2] < \infty$. Let $\underline{x}, \overline{x}$ be constants such that $-\infty \leq \underline{x} < \overline{x} \leq \infty$. Then*

$$\int_{\underline{x}}^{\overline{x}} \omega_X(x)\, dx = \frac{\operatorname{Cov}\left(\max\{\min\{X_t, \overline{x}\}, \underline{x}\}, X_t\right)}{\operatorname{Var}(X_t)}.$$

Proof. By Fubini's theorem and linearity of the covariance operator,

$$\int_{\underline{x}}^{\overline{x}} \operatorname{Cov}(\mathbb{1}\{X_t \geq x\}, X_t)\, dx = \operatorname{Cov}\left(\int_{\underline{x}}^{\overline{x}} \mathbb{1}\{X_t \geq x\}\, dx, X_t\right).$$

Considering separately the three cases $X_t < \underline{x}$, $X_t \in [\underline{x}, \overline{x}]$, and $X_t > \overline{x}$, it can be verified that

$$\int_{\underline{x}}^{\overline{x}} \mathbb{1}\{X_t \geq x\}\, dx = \max\{\min\{X_t, \overline{x}\}, \underline{x}\} - \underline{x}. \qquad \square$$

Note that the lemma holds even if X_t has a discrete distribution (e.g., the empirical distribution). It implies in particular that the ordinary least squares (OLS)-estimated weight function discussed in Section 3.1 integrates to 1 across all $x \in \mathbb{R}$ in finite samples.

A.2 Identification with instruments under endogeneity

We now generalize the setup in Section 3.2 by allowing X_t to be endogenous and incorporating covariates $\mathbf{W}_t$. In particular, we retain the nonparametric structural model (1), but drop the independence assumption (2) and the selection-on-observables assumption (17). Let

$$X_t = \xi(Z_t, \mathbf{W}_t, \tilde{V}_t) \tag{A.1}$$

denote the first-stage equation, with $\tilde{V}_t$ corresponding to the unobservable determinants of X_t. To accommodate a variety of alternatives to the classic Imbens and Angrist (1994) monotonicity assumption, we follow Small, Tan, Ramsahai, Lorch, and Brookhart (2017) and suppose there is a vector $\mathbf{V}_t$ (not necessarily observable) containing the covariates $\mathbf{W}_t$ that is a sufficient statistic for endogeneity in the sense that

$$E[\psi_h(x, \mathbf{U}_{h,t+h}) \mid X_t, Z_t, \mathbf{V}_t] = \Psi_h(x, \mathbf{V}_t), \tag{A.2}$$

where

$$\Psi_h(x, \mathbf{v}) \equiv E[\psi_h(x, \mathbf{U}_{h,t+h}) \mid \mathbf{V}_t = \mathbf{v}]$$

denotes the *marginal treatment response function*. We also make the exclusion restriction that the expectation of Z_t conditional on $\mathbf{V}_t$ depends only on $\mathbf{W}_t$. To ensure it is sufficient to control for the covariates linearly, we further assume the conditional expectation is linear:

$$E[Z_t \mid \mathbf{V}_t] = \mathbf{W}_t' \boldsymbol{\gamma}. \tag{A.3}$$

We assume implicitly that the conditional expectations in equations (A.2) and (A.3) are well-defined.

This setup accommodates several scenarios. On the one hand, including more variables in $\mathbf{V}_t$ makes equation (A.2) less restrictive; on the other hand, as we will see shortly, it requires stronger conditions to ensure non-negative weights in the instrumental variables estimand. As a leading case, we may put $\mathbf{V}_t = (\mathbf{W}_t, \tilde{V}_t)$. Then equations (A.2) and (A.3) hold if the instrument is conditionally randomly assigned (and $E[Z_t \mid \mathbf{W}_t]$ is linear). In this case, with a binary shock X_t, the difference $\Psi_h(1, \mathbf{v}) - \Psi_h(0, \mathbf{v})$ corresponds to the marginal treatment effect of Heckman and Vytlacil (1999, 2005). Second, under selection on observables, we can simply put $\mathbf{V}_t = \mathbf{W}_t$. In this case, equation (A.2) states that Z_t is a valid proxy for X_t, analogously to equation (13); moreover, Ψ_h equals the conditional average structural function. Third, we may put $\mathbf{V}_t = (\mathbf{W}_t, \mathbf{U}_{h,t+h})$, in which case equation (A.2) holds trivially, and $\Psi_h = \psi_h$. See Small, Tan, Ramsahai, Lorch, and Brookhart (2017, Section 6) for examples of other choices for $\mathbf{V}_t$ (which is denoted by $\mathbf{U}$ in their notation) when X_t is assumed to be binary.

Under equations (1), (A.2), and (A.3), it follows by the Frisch-Waugh theorem and iterated expectations that the coefficient on Z_t in a linear "reduced-form" regression of Y_{t+h} onto Z_t and a vector of controls $\mathbf{W}_t$ is given by

$$\tilde{\beta}_h = \frac{E[(Z_t - \mathbf{W}_t'\boldsymbol{\gamma})\Psi_h(X_t, \mathbf{V}_t)]}{E[(Z_t - \mathbf{W}_t'\boldsymbol{\gamma})^2]}. \tag{A.4}$$

As in Section 6.2, we assume that the support of X_t conditional on $\mathbf{V}_t$ is contained in an interval $I_{\mathbf{V}_t}$. If there are gaps in the support of X_t, such as when X_t is discrete, we assume that we can extend $\Psi_h(\cdot, \mathbf{V}_t)$ to $I_{\mathbf{V}_t}$ such that the extension is locally absolutely continuous. Applying Lemma 2 with $\mathbf{V}_t$ playing the role of the covariates $\mathbf{W}$ then yields the following result:

Proposition A.1. *Suppose equation (A.3) holds and that $E[(Z_t - \mathbf{W}_t'\boldsymbol{\gamma})^2]$ is positive and finite. Let $\alpha(\mathbf{V}_t, X_t) = E[Z_t - \mathbf{W}_t'\boldsymbol{\gamma} \mid \mathbf{V}_t, X_t]$. Suppose also that conditional on $\mathbf{V}_t$, the following holds almost surely: (i) the support of X_t is contained in a (possibly unbounded) interval $I_{\mathbf{V}_t} \subseteq \mathbb{R}$; and (ii) $\Psi_h(\cdot, \mathbf{V}_t)$ is locally absolutely continuous on $I_{\mathbf{V}_t}$. Suppose also that (iii) there exists a function $x_0(\mathbf{V}_t) \in I_{\mathbf{V}_t}$ such that $E[|\alpha(\mathbf{V}_t, X_t) \int_{x_0(\mathbf{V}_t)}^{X_t} |\Psi_h'(x, \mathbf{V}_t)|\, dx|] < \infty$; and that (iv) $E[|\alpha(\mathbf{V}_t, X_t)|(1 + |\Psi_h(X_t, \mathbf{V}_t)|)] < \infty$. Then the estimand (A.4) satisfies*

$$\tilde{\beta}_h = E\left[\int \omega(x, \mathbf{V}_t)\Psi_h'(x, \mathbf{V}_t)\, dx\right],$$

where $\omega(x, \mathbf{v}) \equiv E[\mathbb{1}\{X_t \geq x\}(Z_t - \mathbf{W}_t'\boldsymbol{\gamma}) \mid \mathbf{V}_t = \mathbf{v}]/\operatorname{Var}(Z_t - \mathbf{W}_t'\boldsymbol{\gamma})$, and $\Psi_h'(x, \mathbf{v})$ denotes the partial derivative with respect to x.

Proposition A.1 shows that the reduced-form regression of Y_t onto Z_t identifies a weighted average of derivatives of the marginal treatment response function.[1] A sufficient condition ensuring non-negative weights $\omega(x, \mathbf{v})$ is the stochastic monotonicity condition that $E[Z_t \mid X_t = x, \mathbf{V}_t]$ is almost surely monotone increasing (or decreasing) in x.

Applying the result with $\mathbf{V}_t = (\mathbf{W}_t, \tilde{V}_t)$ generalizes Theorem 1 of Angrist, Graddy, and

[1]An analogous application of Lemma 2 to a linear "first-stage regression" of X_t onto Z_t and a vector of controls $\mathbf{W}_t$ shows that it identifies the integral of the weights, $E[\int \omega(x, \mathbf{V}_t)dx]$, so that the weights in the associated instrumental variables regression integrate to one.

Imbens (2000) in several ways: we don't require differentiability of the potential outcome function ψ_h, only of the marginal treatment response function; X_t is not required to be continuous—it may be discrete or mixed; Z_t is not restricted to be binary; we impose no structure on the first-stage equation; and finally, we impose only very weak moment conditions. In this case, the stochastic monotonicity assumption is equivalent to the first-stage monotonicity condition that $\xi(z, \mathbf{W}_t, \tilde{V}_t)$ is increasing in z: this corresponds to Assumption 4 in Angrist, Graddy, and Imbens (2000) if z is binary. Under this condition, $\tilde{\beta}_h$ can be interpreted as identifying a weighted average of marginal effects for compliers. Let $\mathcal{C}_{\mathbf{w}}$ collect all $\tilde{v}$ in the support of $\tilde{V}_t$ such that $\xi(\cdot, \mathbf{w}, \tilde{v})$ is not constant. Following Angrist, Imbens, and Rubin (1996), we refer to the set $\mathcal{C}_{\mathbf{w}}$ as the set of *compliers*, since if $\tilde{V}_t \in \mathcal{C}_{\mathbf{W}_t}$, the shock X_t complies with the instrument assignment in the sense that it increases with Z_t. If $\tilde{V}_t \notin \mathcal{C}_{\mathbf{W}_t}$, variation in the instrument Z_t has no impact on X_t, and hence $\omega(x, \mathbf{W}_t, \tilde{V}_t) = 0$ for all x. Thus, the estimand $\tilde{\beta}_h$ only places positive weight on the marginal effect $\Psi'_h(x, \mathbf{W}_t, \tilde{V}_t)$ for compliers.

As discussed in Small, Tan, Ramsahai, Lorch, and Brookhart (2017) in the context with a binary treatment, the first-stage monotonicity assumption may be too strong in some contexts. In such scenarios, other choices of $\mathbf{V}_t$ may be preferable, such as setting $\mathbf{V}_t = (\mathbf{W}_t, \mathbf{U}_{h,t+h})$. For this choice of $\mathbf{V}_t$, Proposition A.1 generalizes Proposition 1 in Borusyak and Hull (2024) by allowing X_t to have full support, and dropping the requirement that X_t be continuous.

If X_t is exogenous, we may set $\mathbf{V}_t = \mathbf{W}_t$, which both weakens the condition ensuring non-negative weights and broadens the interpretation of the estimand. In particular, now $\Psi'_h(x, \mathbf{v}) = \partial E[\psi_h(x, \mathbf{U}_{h,t+h}) \mid \mathbf{W}_t = \mathbf{w}]/\partial x$ gives the overall marginal effect, not just the effect for compliers. Also, now the stochastic monotonicity condition requires only that $E[Z_t \mid X_t = x, \mathbf{W}_t]$ is increasing in x. Without covariates, this reduces to the condition that $\zeta(x) = E[Z_t \mid X_t = x]$ is monotone, as discussed in Section 3.2. This is clearly weaker than the Angrist, Graddy, and Imbens (2000) condition that $E[Z_t \mid X_t = x, \mathbf{W}_t, \tilde{V}_t]$ is increasing in x: we only require this to be true *on average* over $\tilde{V}_t$ rather than for almost all realizations of $\tilde{V}_t$. This condition holds for many measurement error models for Z_t, even though the stronger first-stage monotonicity condition may be violated.

A.3 Identification via heteroskedasticity: linear case

Here we derive the linear identification result (23), following Rigobon and Sack (2004) and Lewbel (2012). Note first that

$$
\begin{aligned}
E[Z \mid \mathbf{U}] &= E[(\theta_1 X + \gamma_1(\mathbf{U}))(D - E[D]) \mid \mathbf{U}] \\
&= \theta_1 E[X(D - E[D]) \mid \mathbf{U}] + \gamma_1(\mathbf{U}) E[D - E(D) \mid \mathbf{U}] \\
&= \theta_1 \operatorname{Cov}(X, D) + \gamma_1(\mathbf{U}) E[D - E(D)] \\
&= 0.
\end{aligned}
$$

Hence,

$$
\operatorname{Cov}(\mathbf{Y}, Z) = \boldsymbol{\theta} \operatorname{Cov}(X, Z) + \operatorname{Cov}(\boldsymbol{\gamma}(\mathbf{U}), Z) = \boldsymbol{\theta} \operatorname{Cov}(X, Z),
$$

and the claim (23) follows, provided that $\operatorname{Cov}(X, Z) \neq 0$. The latter holds if $\theta_1 \neq 0$ and $\operatorname{Cov}(X^2, D) \neq 0$, since

$$
\begin{aligned}
\operatorname{Cov}(X, Z) &= E[X(\theta_1 X + \gamma_1(\mathbf{U}))(D - E[D])] \\
&= \theta_1 \operatorname{Cov}(X^2, D) + \operatorname{Cov}(X, D) E[\gamma_1(\mathbf{U})] \\
&= \theta_1 \operatorname{Cov}(X^2, D). \qquad \square
\end{aligned}
$$

A.4 Details for Example 4

Let $\tilde{U}_1$ and $\tilde{U}_2$ be independent uniforms on $[0, 1]$. By the Box-Muller transform, the two variables

$$
\tilde{Y}_1 \equiv \sqrt{-2 \log \tilde{U}_1} \cos(2\pi \tilde{U}_2), \quad \tilde{Y}_2 \equiv \sqrt{-2 \log \tilde{U}_1} \sin(2\pi \tilde{U}_2),
$$

have a bivariate standard normal distribution.

Define $X \equiv \log(-2 \log \tilde{U}_1)$ and $U \equiv \log \cos^2(2\pi \tilde{U}_2)$, so that X and U are independent and non-Gaussian. By construction, the following two variables are independent:

$$
Y_1 \equiv \log \tilde{Y}_1^2 = X + U, \quad Y_2 \equiv \log \tilde{Y}_2^2 = X + \gamma(U),
$$

where

$$
\gamma(u) \equiv \log\left(1 - \exp(u)\right), \quad u < 0,
$$

and we have used that $\exp(U) = \cos^2(2\pi \tilde{U}_2) = 1 - \sin^2(2\pi \tilde{U}_2)$. Note that in this example, the shocks X and U do not have mean zero as commonly assumed in the literature, but this is easily rectified by just subtracting off their means in the calculations. □

A.5 Additional empirical estimates of causal weights

Complementing the results for government spending shocks in Figure 1 (Section 3.1), Figures A.1 to A.3 show estimated causal weight functions for several identified tax shocks, technology shocks, and monetary policy shocks. The data is obtained from the replication files for Ramey (2016), as discussed in Section 3.1. While many of the shocks yield approximately symmetric weight functions, the Romer and Romer (2010) and Mertens and Ravn (2014) tax shocks are both skewed towards tax cuts, while the Christiano, Eichenbaum, and Evans (1999) and Gertler and Karadi (2015) monetary shocks are skewed towards interest rate cuts. As discussed in Section 3.1, this is important to keep in mind when using impulse response estimates to discipline structural models that feature asymmetries.

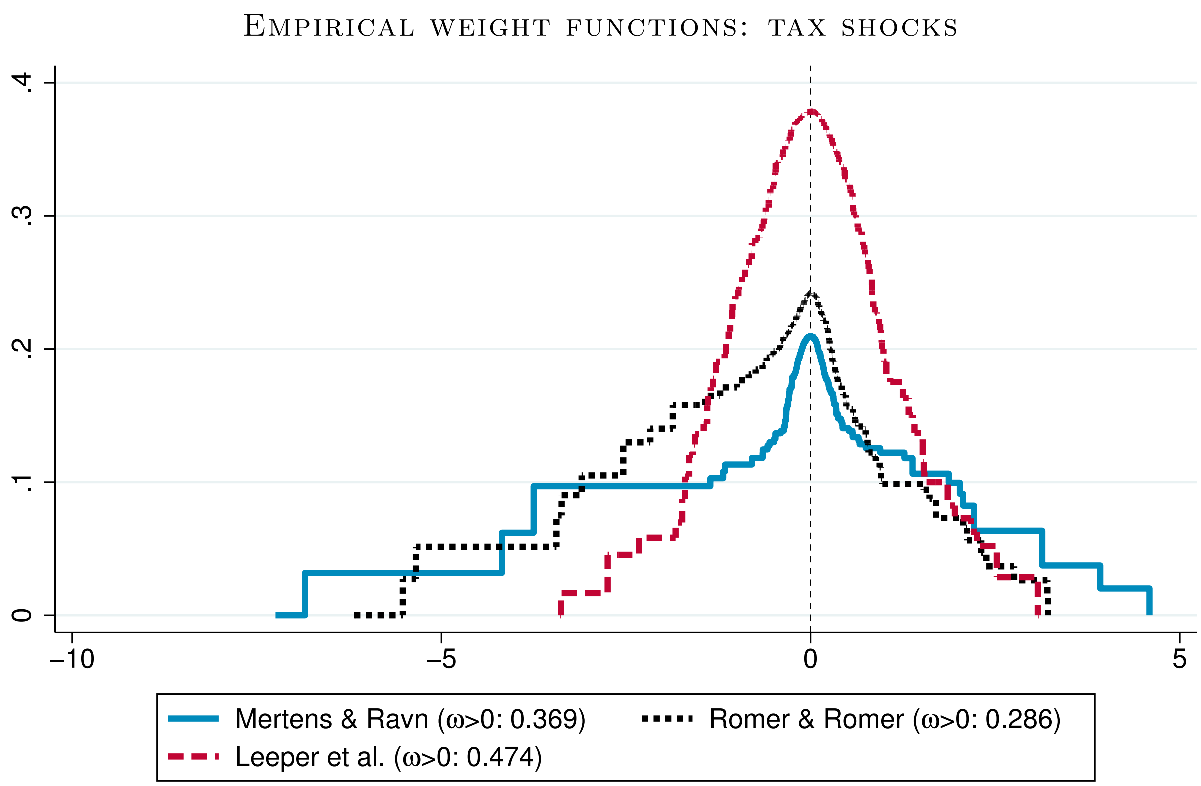

Figure A.1: Estimated causal weight functions ω_X for tax shocks obtained from the replication files for Ramey (2016), quarterly data. Horizontal axis in units of standard deviations. "$\omega > 0$": total weight $\int_0^\infty \omega_X(x)\,dx$ on positive shocks. Papers referenced: Mertens and Ravn (2014), Romer and Romer (2010), Leeper, Richter, and Walker (2012).

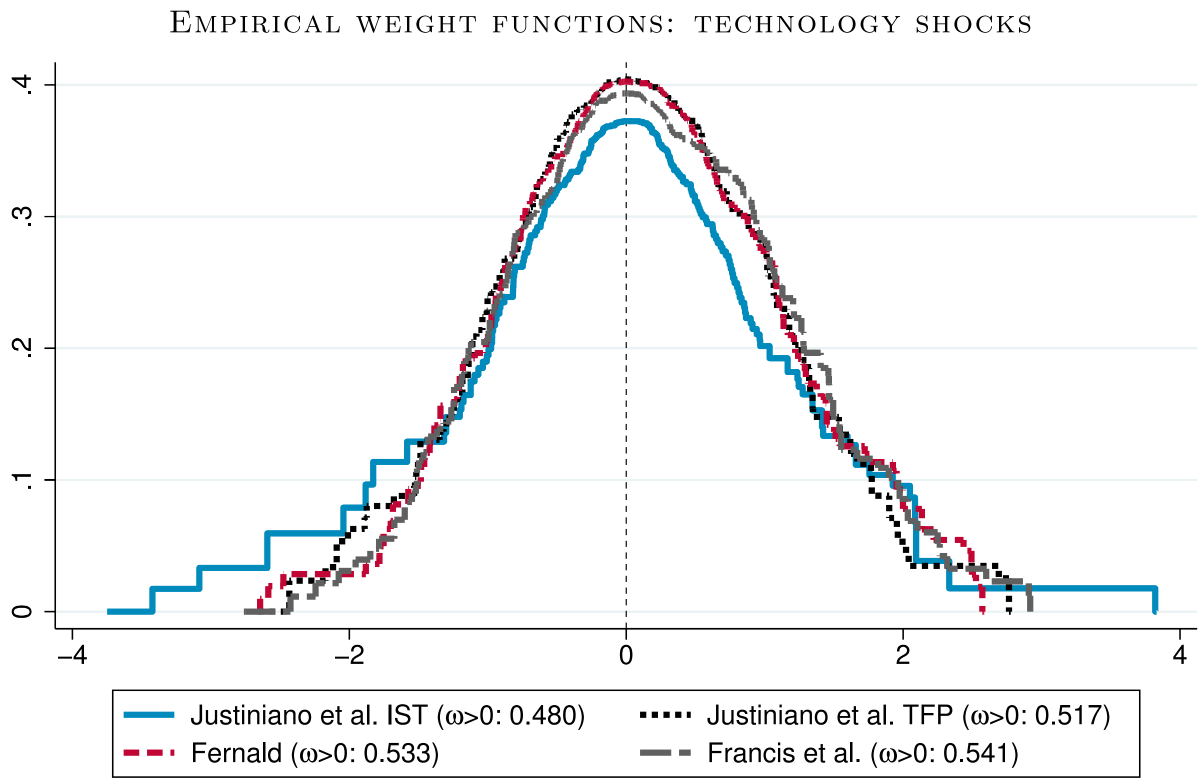

Figure A.2: Estimated causal weight functions ω_X for technology shocks obtained from the replication files for Ramey (2016), quarterly data. Horizontal axis in units of standard deviations. "TFP" = total factor productivity. "IST" = investment-specific technology. "$\omega > 0$": total weight $\int_0^\infty \omega_X(x)\,dx$ on positive shocks. Papers referenced: Justiniano, Primiceri, and Tambalotti (2011), Fernald (2014), Francis, Owyang, Roush, and DiCecio (2014).

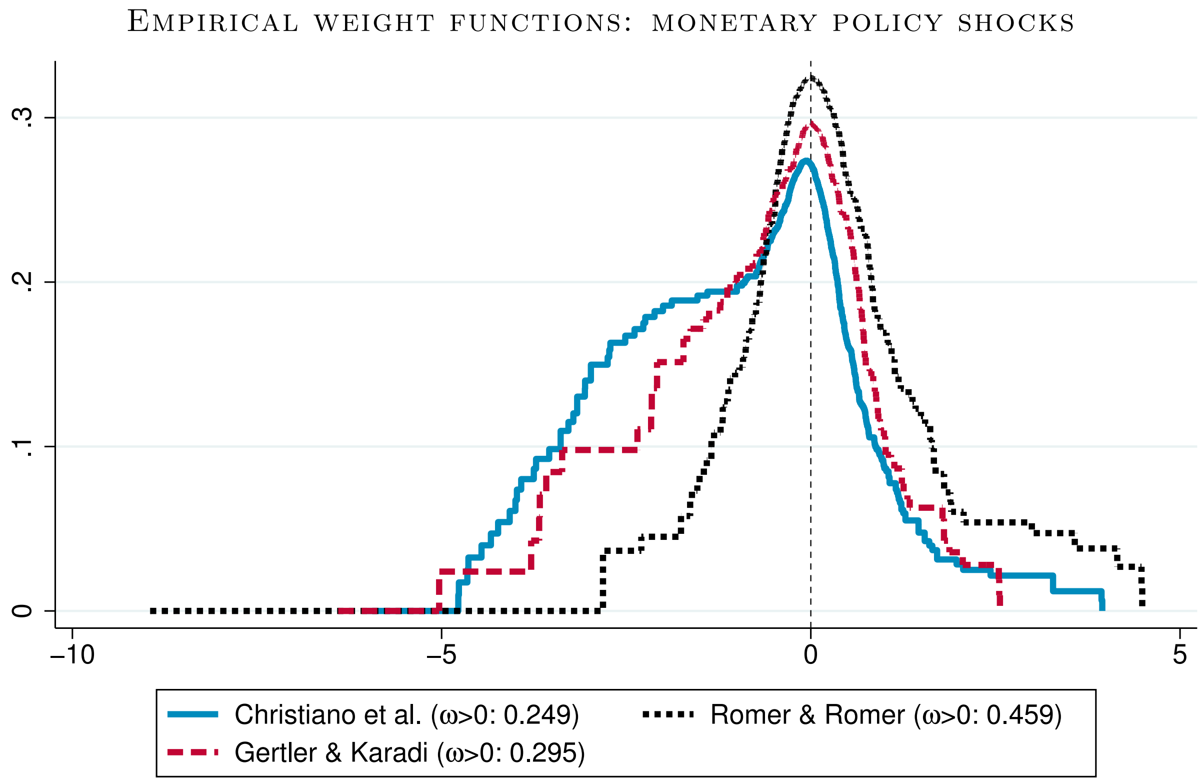

Figure A.3: Estimated causal weight functions ω_X for monetary policy shocks obtained from the replication files for Ramey (2016), quarterly data. Horizontal axis in units of standard deviations. "$\omega > 0$": total weight $\int_0^\infty \omega_X(x)\,dx$ on positive shocks. Papers referenced: Christiano, Eichenbaum, and Evans (1999), Romer and Romer (2010), Gertler and Karadi (2015).

Appendix B Proofs

B.1 Auxiliary lemma

Lemma B.1. *Suppose that conditions (i)–(iii) of Lemma 1 hold. Suppose additionally that for some* $\underline{x}, \overline{x} \in I$, $\underline{x} \leq \overline{x}$, *it holds that either (a)* $\alpha(x)$ *only changes sign for* $x \in [\underline{x}, \overline{x}]$ *and* $\int_I |\omega(x)g'(x)|\,dx < \infty$, *or (b)* $g(x)$ *is monotone for* $x \leq \underline{x}$ *and for* $x \geq \overline{x}$. *Then condition (iv) of Lemma 1 holds for any* $x_0 \in [\underline{x}, \overline{x}]$.

Proof. Bound

$$E\left[\left|\alpha(X)\int_{x_0}^{X}|g'(x)|\,dx\right|\right] \leq E[|\alpha(X)|]\int_{\underline{x}}^{\overline{x}}|g'(x)|\,dx + E\left[\mathbb{1}\{X \geq \overline{x}\}|\alpha(X)|\int_{\overline{x}}^{X}|g'(x)|\,dx\right] + E\left[\mathbb{1}\{X \leq \underline{x}\}|\alpha(X)|\int_{X}^{\underline{x}}|g'(x)|\,dx\right].$$

The first term on the right-hand side is finite since g is absolutely continuous on $[\underline{x}, \overline{x}]$. Now consider the second term on the right-hand side; the third term can be handled analogously. Under condition (a), $\alpha(x)$ has the same sign for all $x \geq \overline{x}$, so the second term equals

$$\int_I \mathbb{1}\{x \geq \overline{x}\}\,|E[\mathbb{1}\{X \geq x\}\alpha(X)]|\,|g'(x)|\,dx \leq \int_I |\omega(x)||g'(x)|\,dx < \infty.$$

Under condition (b), since $g(x)$ is monotone for $x \geq \overline{x}$, $\int_{\overline{x}}^{X}|g'(x)|\,dx = |\int_{\overline{x}}^{X} g'(x)\,dx|$, so that the second term on the right-hand side in the first display equals

$$E\left[\mathbb{1}\{X \geq \overline{x}\}|\alpha(X)|\left|\int_{\overline{x}}^{X} g'(x)\,dx\right|\right] = E\left[\mathbb{1}\{X \geq \overline{x}\}|\alpha(X)||g(X) - g(\overline{x})|\right]$$
$$\leq E[|\alpha(X)g(X)|] + |g(\overline{x})|E[|\alpha(X)|] < \infty. \qquad \square$$

B.2 Proof of Proposition 1

This is a special case of Proposition 3 with $Z = \zeta(X) = X$. Lemma A.1 implies that the weights integrate to 1. $\square$

B.3 Proof of Proposition 2

Since $g_h'(x)$ is locally absolutely continuous and $E[|g_h''(X_t)|] < \infty$, by Stein's lemma (Stein, 1981, Lemma 1),

$$E[g_h''(X_t)] = E[X_t g_h'(X_t)].$$

Since $E[|g_h(X_t)|] < \infty$, another application of Stein's lemma yields $E[X_t g_h'(X_t) + g_h(X_t)] = E[X_t^2 g_h(X_t)]$. Hence, $\operatorname{Cov}(g_h(X_t), X_t^2) = E[X_t g_h'(X_t)] = E[g_h''(X_t)]$. A third application of Stein's lemma yields $E[g_h'(X_t)] = \operatorname{Cov}(X_t, g_h(X_t))$. The result then follows from the definitions (10)–(11). $\square$

B.4 Proof of Proposition 3

The representation of the estimand follows directly from Lemma 1 and Lemma B.1 with $\alpha(X_t) = \zeta(X_t) - E[Z_t]$. Claim (i) for the weights follows from a simple calculation. Claim (ii) follows from $\operatorname{Cov}(\mathbb{1}\{X_t \geq x\}, \zeta(X_t)) = \operatorname{Var}(\mathbb{1}\{X_t \geq x\})\{E[\zeta(X_t) \mid X_t \geq x] - E[\zeta(X_t) \mid X_t < x]\}$. For the last statement of the proposition, observe that for $x_U > x_L$, $\tilde{\omega}_Z(x_L) - \tilde{\omega}_Z(x_U)$ is proportional to $E[\mathbb{1}\{x_L < X_t < x_U\}(\zeta(X_t) - E[Z_t])]$, which is positive if $x_0 < x_L < x_U$ and negative if $x_L < x_U < x_0$. $\square$

B.5 Proof of Proposition 4

Let τ be a Rademacher random variable independent of $(D, R, \mathbf{U})$, i.e., $P(\tau = 1 \mid D, R, \mathbf{U}) = P(\tau = -1 \mid D, R, \mathbf{U}) = 1/2$. Since the distribution of R is symmetric around zero, R has the same distribution as $|R| \times \tau$, and thus $(X, \mathbf{U})$ has the same distribution as $(|X|\tau, \mathbf{U})$. Let $\tilde{U}$ be uniform on $[0,1]$ independently of (D, R), and let $\phi_\tau \colon \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{\phi}_{\mathbf{U}} \colon \mathbb{R} \to \mathbb{R}^{m-1}$ be measurable functions such that $(\tau, \mathbf{U})$ has the same distribution as $(\phi_\tau(\tilde{U}), \boldsymbol{\phi}_{\mathbf{U}}(\tilde{U}))$ (see the discussion after Proposition 6 on the construction of such functions). Then it follows that $(X, \mathbf{U})$ has the same distribution as $(|X|\phi_\tau(\tilde{U}), \boldsymbol{\phi}_{\mathbf{U}}(\tilde{U}))$, and the conclusion of the proposition obtains by defining $\tilde{\boldsymbol{\psi}}(x, \tilde{u}) \equiv \boldsymbol{\psi}(|x|\phi_\tau(\tilde{u}), \boldsymbol{\phi}_{\mathbf{U}}(\tilde{u}))$. $\square$

B.6 Proof of Proposition 5

Since $\boldsymbol{\gamma}(\mathbf{U})$ is independent of (X, D) with mean zero,

$$\begin{aligned}\operatorname{Cov}(\mathbf{Y}, Z \mid X) &= \operatorname{Cov}(\boldsymbol{\gamma}(\mathbf{U}), (\Psi_1(X) + \gamma_1(\mathbf{U}))(D - E[D]) \mid X) \\ &= \operatorname{Cov}(\boldsymbol{\gamma}(\mathbf{U}), \gamma_1(\mathbf{U}))\{E[D \mid X] - E[D]\}.\end{aligned}$$

The law of total covariance therefore implies

$$\begin{aligned}\mathrm{Cov}(\mathbf{Y}, Z) &= E[\mathrm{Cov}(\mathbf{Y}, Z \mid X)] + \mathrm{Cov}(E[\mathbf{Y} \mid X], E[Z \mid X]) \\ &= 0 + E[\boldsymbol{\Psi}(X)\{E[Z \mid X] - E[Z]\}].\end{aligned}$$

The result now follows from Lemma 1 and Lemma B.1, with weights given by

$$\begin{aligned}\check{\omega}(x) &\equiv E[\mathbb{1}\{X \geq x\}\{E[Z \mid X] - E[Z]\}] \\ &= \mathrm{Cov}(\mathbb{1}\{X \geq x\}, E[Z \mid X, D]) \\ &= \mathrm{Cov}(\mathbb{1}\{X \geq x\}, \Psi_1(X)(D - E[D])),\end{aligned}$$

where the last equality follows from

$$E[Z \mid X, D] = E[Y_1 \mid X, D](D - E[D]) = \Psi_1(X)(D - E[D]). \qquad \square$$

B.7 Proof of Proposition 6

Let $Q_j(\tau \mid \tilde{Y}_{j-1}, \tilde{Y}_{j-2}, \ldots, \tilde{Y}_1)$ denote the τ-th quantile of $\tilde{Y}_j$ conditional on $\tilde{Y}_{j-1}, \tilde{Y}_{j-2}, \ldots, \tilde{Y}_1$. Now construct an n-dimensional vector $\mathbf{Y}^* = (Y_1^*, \ldots, Y_n^*)$ as follows. First set $Y_1^* \equiv \tilde{Y}_1$. Then for $j > 1$, let $Y_j^* \equiv Q_j(\bar{U}_{j-1} \mid \tilde{Y}_{j-1} = Y_{j-1}^*, \ldots, \tilde{Y}_1 = Y_1^*)$. Standard arguments yield that $\mathbf{Y}^*$ has the same distribution as $\tilde{\mathbf{Y}}$. Consequently, $\bar{\mathbf{Y}} \equiv \boldsymbol{\Upsilon}^{-1}(\mathbf{Y}^*)$ has the same distribution as $\mathbf{Y} = \boldsymbol{\Upsilon}^{-1}(\tilde{\mathbf{Y}})$. The mapping from $(\tilde{X}, \bar{U}_1, \ldots, \bar{U}_{n-1})$ to $\mathbf{Y}^*$ is continuous by the assumptions on Q_j, and so is the implied $\bar{\boldsymbol{\psi}}$ mapping by continuity of $\boldsymbol{\Upsilon}^{-1}$. $\square$

B.8 Proof of Lemma 1

This result follows directly from Lemma 2 by letting $\mathbf{W}$ equal a constant. $\square$

B.9 Proof of Lemma 2

Observe

$$\begin{aligned}E\left[\int \omega(x, \mathbf{W}) g'(x, \mathbf{W})\, dx\right] &= E\left[\int_{I_{\mathbf{W}}} E[\mathbb{1}\{X \geq x \geq x_0(\mathbf{W})\}\alpha(X, \mathbf{W}) \mid \mathbf{W}] g'(x, \mathbf{W})\, dx\right] \\ &\quad - E\left[\int_{I_{\mathbf{W}}} E[\mathbb{1}\{X < x < x_0(\mathbf{W})\}\alpha(X, \mathbf{W}) \mid \mathbf{W}] g'(x, \mathbf{W})\, dx\right] \\ &= E\left[\int_{I_{\mathbf{W}}} \mathbb{1}\{X \geq x \geq x_0(\mathbf{W})\}\alpha(X, \mathbf{W}) g'(x, \mathbf{W})\, dx\right]\end{aligned}$$

$$
\begin{aligned}
&\quad - E\left[\int_{I_{\mathbf{W}}} \mathbb{1}\{X < x < x_0(\mathbf{W})\}\alpha(X, \mathbf{W})g'(x, \mathbf{W})\, dx\right] \\
&= E\left[\mathbb{1}\{X \geq x_0(\mathbf{W})\}\alpha(X, \mathbf{W})(g(X, \mathbf{W}) - g(x_0(\mathbf{W}), \mathbf{W}))\right] \\
&\quad - E\left[\mathbb{1}\{X < x_0(\mathbf{W})\}\alpha(X, \mathbf{W})(g(x_0(\mathbf{W}), \mathbf{W}) - g(X, \mathbf{W}))\right] \\
&= E\left[\alpha(X, \mathbf{W})(g(X, \mathbf{W}) - g(x_0(\mathbf{W}), \mathbf{W}))\right] \\
&= E\left[\alpha(X, \mathbf{W})g(X, \mathbf{W})\right],
\end{aligned}
$$

where the first equality uses the fact that since $E[\alpha(X, \mathbf{W}) \mid \mathbf{W}] = 0$ by condition (iii), $\omega(x, \mathbf{w}) = -E[\mathbb{1}\{X < x\}\alpha(X, \mathbf{w}) \mid \mathbf{W} = \mathbf{w}]$, the second equality uses Fubini's theorem, which is justified since both integrals exist by condition (iv), the third equality follows by the fundamental theorem of calculus and condition (ii), the fourth equality collects terms, and the last equality uses iterated expectations, which is justified since

$$
\begin{aligned}
&E\left[|\alpha(X, \mathbf{W})g(x_0(\mathbf{W}), \mathbf{W})|\right] \\
&\leq E\left[|\alpha(X, \mathbf{W})g(X, \mathbf{W})|\right] + E\left[|\alpha(X, \mathbf{W})(g(X, \mathbf{W}) - g(x_0(\mathbf{W}), \mathbf{W}))|\right] \\
&\leq E\left[|\alpha(X, \mathbf{W})g(X, \mathbf{W})|\right] + E\left[\left|\alpha(X, \mathbf{W}) \int_{x_0(\mathbf{W})}^{X} |g'(x, \mathbf{W})|\, dx\right|\right] < \infty,
\end{aligned}
$$

by conditions (iv) and (v). □

B.10 Proof of Proposition 7

Observe that under either condition (a) or condition (b),

$$
\begin{aligned}
&E[(X - \pi(\mathbf{W}))g(X, \mathbf{W})] \\
&\qquad = E[(X - \pi^*(\mathbf{W}))g(X, \mathbf{W})] + E\left[(\pi^*(\mathbf{W}) - \pi(\mathbf{W})) \int \lambda(x, \mathbf{W})g'(x, \mathbf{W})\, dx\right].
\end{aligned}
$$

Applying Lemma 2 with $\alpha(X, \mathbf{W}) = X - \pi^*(\mathbf{W})$ and $x_0(\mathbf{W}) = \pi^*(\mathbf{W})$ yields

$$
E[(X - \pi^*(\mathbf{W}))g(X, \mathbf{W})] = E\left[\int \omega^*(x, \mathbf{W})g'(x, \mathbf{W})\, dx\right].
$$

Note that condition (iv) of Lemma 2 follows from a similar argument as in Lemma B.1 (conditional on $\mathbf{W}$).

Since $(X - \pi(\mathbf{W}))$ is orthogonal to $\pi(\mathbf{W})$ and to a constant function,

$$\begin{aligned}\operatorname{Var}(X - \pi(\mathbf{W})) &= E[(X - \pi^*(\mathbf{W}))X] + E[(\pi^*(\mathbf{W}) - \pi(\mathbf{W}))\pi^*(\mathbf{W})] \\ &= E[(X - \pi^*(\mathbf{W}))X] + E\left[(\pi^*(\mathbf{W}) - \pi(\mathbf{W})) \int \lambda(x, \mathbf{W})\, dx\right],\end{aligned}$$

and it follows that the weights integrate to one. The last statement of the proposition can be shown using the same argument as in the proof of Proposition 3. □

References

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434), 444–455.

Borusyak, K., and P. Hull (2024): "Negative Weights Are No Concern in Design-Based Specifications," *AEA Papers and Proceedings*, 114, 597–600.

Christiano, L., M. Eichenbaum, and C. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1A, pp. 65–148. Elsevier.

Fernald, J. G. (2014): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Working Paper 2012–19, Federal Reserve Bank of San Francisco.

Francis, N., M. T. Owyang, J. E. Roush, and R. DiCecio (2014): "A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks," *The Review of Economics and Statistics*, 96(4), 638–647.

Gertler, M., and P. Karadi (2015): "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7(1), 44–76.

Heckman, J. J., and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects.," *Proceedings of the National Academy of Sciences*, 96(8), 4730–4734.

——— (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2011): "Investment shocks and the relative price of investment," *Review of Economic Dynamics*, 14(1), 102–121, Special issue: "Sources of Business Cycles".

Leeper, E. M., A. W. Richter, and T. B. Walker (2012): "Quantitative Effects of Fiscal Foresight," *American Economic Journal: Economic Policy*, 4(2), 115–44.

Lewbel, A. (2012): "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics*, 30(1), 67–80.

Mertens, K., and M. O. Ravn (2014): "A reconciliation of SVAR and narrative estimates of tax multipliers," *Journal of Monetary Economics*, 68, S1–S19, Supplement issue: October 19-20, 2012 Research conference on "Financial Markets, Financial Policy, and Macroeconomic Activity" sponsored by the Study Center Gerzensee and the Swiss National Bank.

Ramey, V. A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 2, pp. 71–162. Elsevier.

Rigobon, R., and B. Sack (2004): "The impact of monetary policy on asset prices," *Journal of Monetary Economics*, 51(8), 1553–1575.

Romer, C. D., and D. H. Romer (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100(3), 763–801.

Small, D. S., Z. Tan, R. R. Ramsahai, S. A. Lorch, and M. A. Brookhart (2017): "Instrumental Variable Estimation with a Stochastic Monotonicity Assumption," *Statistical Science*, 32(4), 561–579.

Stein, C. M. (1981): "Estimation of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 9(6), 1135–1151.