a paper by **Michal Kolesár** & **Mikkel Plagborg-Møller (Princeton)**

Dynamic Causal Effects in a Nonlinear World

The Good, The Bad & The Ugly

January 4, 2025

Impulse responses in a nonlinear world

- Impulse response: dynamic causal effect of shock (policy/fundamental) on outcome.
- Macro modelers and policy-makers think nonlinearities are essential.
  - Thresholds/regimes, occasionally binding constraints, kinks, asymmetries...
- ...but the most popular impulse response estimators are motivated by linear models: SVAR, local projection. Are they useful in a nonlinear world?

This paper

1. Good news: LP/VAR on observed shock/proxy delivers positively weighted avg of causal effects regardless of nonlinearity.
    - Weight function easily estimable by regression. We give empirical examples.
    - In contrast, nonlinear estimators can get sign of marginal effects wrong under misspecification.
2. Bad/ugly news: ID of latent shocks via heteroskedasticity or non-Gaussianity highly sensitive to linearity assumption.
    - Lesson: hard work of directly measuring shocks/proxies pays off.
3. Building block: new results on identification of weighted average marginal effects.

Literature

- Average marginal effects and weighted regressions: Yitzhaki (1996); Newey & Stoker (1993); Angrist & Krueger (1999); Angrist (2001); Goldsmith-Pinkham, Hull & Kolesár (2024)
- Semiparametric causal time series: Gallant, Rossi & Tauchen (1993); White (2006); Angrist & Kuersteiner (2011); Angrist, Jordà; Kuersteiner (2018); Kitagawa, Wang & Xu (2023)
- LP under nonlinearity: Rambachan & Shephard (2021); Gonçalves, Herrera, Kilian & Pesavento (2021, 2024); Gouriéroux & Lee (2023); Caravello & Martínez Bruera (2024); Casini & McCloskey (2024)
- Finite-sample properties of LP/VAR: Herbst & Johannsen (2024)

Outline

Nonparametric model

- General nonlinear model for outcome Y given shock of interest X and nuisance shocks $\mathbf{U}$:

$$Y_{t+h} = \psi_h(X_t, \mathbf{U}_{h,t+h}), \quad X_t \perp\!\!\!\perp \mathbf{U}_{h,t+h}.$$

- **Structural function** ψ_h captures all direct and indirect effects of X.
- Assume for now X is cts'ly distributed. General case later (e.g., discrete/mixed).
- Example: In endogenous regime switching AR(1) model

$$Y_t = \rho_{t-1} Y_{t-1} + \tau X_t + \nu_t, \quad \text{with} \quad \rho_{t-1} \equiv \rho_0 + (\rho_1 - \rho_0)\,\mathbb{1}\{X_{t-1} + \xi_{t-1} \leq 0\},$$

ψ_h also takes into account effect of X_t on Y_{t+h} via future regimes $\rho_{t+\ell}$.

Causal effects

$$Y_{t+h} = \psi_h(X_t, \mathbf{U}_{h,t+h}), \quad X_t \perp\!\!\!\perp \mathbf{U}_{h,t+h}$$

- **Average structural function**:

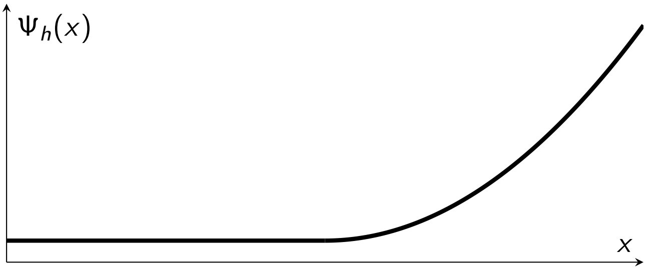
$$\Psi_h(x) \equiv E[\psi_h(x, \mathbf{U}_{h,t+h})], \quad x \in \mathbb{R}.$$

- Object of interest is **average marginal effect**:

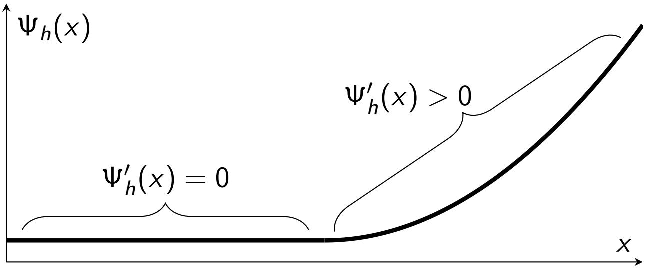
$$\theta_h(\omega) \equiv \int \omega(x) \Psi'_h(x)\, dx.$$

  - Most direct interpretation of $\Psi'_h(x)$: effect of infinitesimal shock $x \mapsto x + \delta$.
  - $\omega(\cdot)$ weights baseline values of shock. Matters in nonlinear models!

Causal effects: graphical example

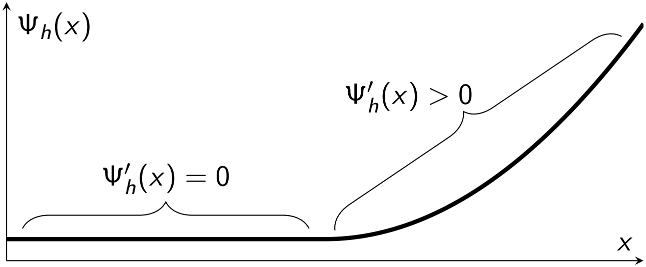


Causal effects: graphical example



- Avg marg'l effect: $\theta_h(\omega) = \int \omega(x)\Psi'_h(x)\,dx$, weighted avg of heterogeneous slopes.

Causal effects: graphical example



- Avg marg'l effect: $\theta_h(\omega) = \int \omega(x)\Psi'_h(x)\,dx$, weighted avg of heterogeneous slopes.
- Nonnegative weights $\omega(\cdot) \geq 0$ desirable, since rules out sign reversal: $\theta_h(\omega) < 0$ despite Ψ_h monotonically increasing. Would be concerning for model calibration/validation.

Outline

1. Nonparametric framework for dynamic causality
2. The Good: observed shocks and proxies
   - Observed shocks
   - Proxies
3. The Bad: identification via heteroskedasticity
4. The Ugly: identification via non-Gaussianity
5. Identification of average marginal effects
6. Conclusion

Outline

1. Nonparametric framework for dynamic causality
2. The Good: observed shocks and proxies
   - Observed shocks
   - Proxies
3. The Bad: identification via heteroskedasticity
4. The Ugly: identification via non-Gaussianity
5. Identification of average marginal effects
6. Conclusion

Identification with observed shocks

- If we observe X, then ASF is identified: Gouriéroux & Lee (2023); Gonçalves et al. (2024)

$$\Psi_h(x) \equiv E[\psi_h(x, \mathbf{U}_{h,t+h})] = E[\psi_h(x, \mathbf{U}_{h,t+h}) \mid X_t = x] = E[Y_{t+h} \mid X_t = x] \equiv g_h(x).$$

But fully nonparametric estimation of g_h is challenging in typical macro data sets.

- Instead, we study what is estimated when running linear local projection

$$Y_{t+h} = \hat{\beta}_h X_t + \hat{\gamma}_h' \mathbf{W}_t + \text{residual}_{h,t+h}.$$

- Assuming X_t is a shock, so $\text{Cov}(X_t, \mathbf{W}_t) = 0$, large-sample **estimand** equals

$$\beta_h \equiv \frac{\text{Cov}(g_h(X_t), X_t)}{\text{Var}(X_t)}.$$

  - SVAR shares exact same estimand given sufficient lags. P-M & Wolf (2021)

Robustness of linear procedures

- Proposition (Yitzhaki, 1996; Rambachan & Shephard, 2021): Linear LP/VAR estimate useful causal summary, regardless of extent of nonlinearities.

$$\beta_h = \int \omega_X(x) g_h'(x)\, dx, \quad \text{where} \quad \omega_X(x) \equiv \frac{\text{Cov}(\mathbb{1}\{X_t \geq x\}, X_t)}{\text{Var}(X_t)}.$$

- Properties of weight function:

  1. Convex: $\omega_X(\cdot) \geq 0$, $\int \omega_X(x)\, dx = 1$.
  2. Hump-shaped: increasing from 0 to its max for $x \leq E[X_t]$, then decreasing back to 0.
  3. Depends only on marginal distribution of X_t, not on $(Y_{t+h} \mid X_t)$ or h.

- Our regularity conditions weaker than literature; just require well-defined β_h and integral.
  - Allow non-smooth structural fct ψ_h, general X distr'n (potentially unbounded support).

Estimating the weight function

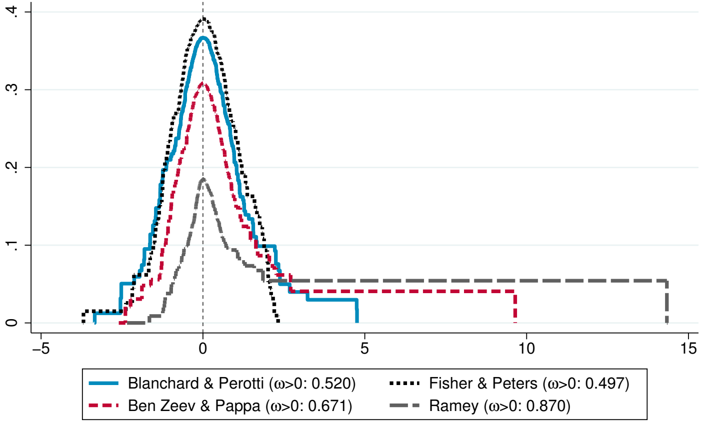
$$\beta_h = \int \omega_X(x) g_h'(x)\, dx, \quad \omega_X(x) \equiv \underbrace{\mathrm{Cov}(\mathbb{1}\{X_t \geq x\}, X_t) / \mathrm{Var}(X_t)}_{\text{regression coefficient}}$$

- $\hat{\omega}_X(x)$: slope in regression of $\mathbb{1}(X_t \geq x)$ on X_t.
- Weights transform empirical CDF of shocks into interpretable units.
    - Visualize asymmetry, outliers, etc.
- Special case: if X_t is Gaussian, then $\omega_X(x) =$ density of X_t. Yitzhaki (1996)

Estimating the weight function

$$\beta_h = \int \omega_X(x) g_h'(x)\, dx, \quad \omega_X(x) \equiv \underbrace{\text{Cov}(\mathbb{1}\{X_t \geq x\}, X_t) / \text{Var}(X_t)}_{\text{regression coefficient}}$$

Government spending shocks from Ramey (2016) handbook chapter:



Sensitivity of nonlinear parametric regression

- If we care about characterizing nonlinearities, we should model them.
- But if we only care about average marginal effects, nonlinear modeling can be counterproductive. Even if variables have limited support, e.g., ZLB! Angrist (2001)
- Illustrative example: (population) LP with quadratic term

$$Y_{t+h} = \beta_{0,h} + \beta_{1,h} X_t + \beta_{2,h} X_t^2 + \text{residual}_{h,t+h},$$

with estimated first derivative $\bar{\beta}_h(x) \equiv \beta_{1,h} + 2\beta_{2,h} x$.

- Proposition: If $X_t \sim N(0,1)$,

$$\bar{\beta}_h(x) = E[(1 + X_t x) g_h'(X_t)] = E[g_h'(X_t)] + x E[g_h''(X_t)].$$

  - Can easily get sign reversals due to negative weights!

Covariates

- Suppose we relax shock independence to "selection on observables":

$$X_t \perp\!\!\!\perp \mathbf{U}_{h,t+h} \mid \mathbf{W}_t.$$

  - Nonparametric version of recursive/Cholesky identification. Angrist & Kuersteiner (2011)

- Then conditional ASF is identified:

$$g_h(x, \mathbf{w}) \equiv E[Y_{t+h} \mid X_t = x, \mathbf{W}_t = \mathbf{w}] = E[\varphi_h(x, \mathbf{U}_{h,t}) \mid \mathbf{W}_t = \mathbf{w}] \equiv \Psi_h(x, \mathbf{w}).$$

- LP with controls

$$Y_{t+h} = \hat{\beta}_h X_t + \hat{\gamma}_h' \mathbf{W}_t + \text{residual}_{h,t+h}$$

  estimates weighted avg of marginal effects $\partial \Psi_h(x, \mathbf{w})/\partial x$.

- But weights need not be positive if "propensity score" $\pi^*(\mathbf{w}) \equiv E[X_t \mid \mathbf{W}_t = \mathbf{w}]$ is nonlinear (more in paper). Lesson: check sensitivity wrt. functional form of controls.

Outline

1. Nonparametric framework for dynamic causality
2. The Good: observed shocks and proxies
   - Observed shocks
   - Proxies
3. The Bad: identification via heteroskedasticity
4. The Ugly: identification via non-Gaussianity
5. Identification of average marginal effects
6. Conclusion

Robustness of linear proxy procedures

- Instead of observing X_t directly, assume we observe a valid **proxy** Z_t satisfying

$$E[Y_{t+h} \mid X_t, Z_t] = E[Y_{t+h} \mid X_t] \equiv g_h(x).$$

- Estimand from "reduced-form" LP/VAR of outcome on proxy:

$$\tilde{\beta}_h \equiv \frac{\mathrm{Cov}(\zeta(X_t), g_h(X_t))}{\mathrm{Var}(Z_t)}, \quad \text{where} \quad \zeta(x) \equiv E[Z_t \mid X_t = x].$$

- Proposition: Proxy identifies weighted sum of marginal effects.

$$\tilde{\beta}_h = \int \tilde{\omega}_Z(x) g_h'(x)\, dx, \quad \text{where} \quad \tilde{\omega}_Z(x) \equiv \frac{\mathrm{Cov}(\mathbb{1}\{X_t \geq x\}, \zeta(X_t))}{\mathrm{Var}(Z_t)}.$$

  - Weights depend only on distr'n of (X_t, Z_t), not $(Y_{t+h} \mid X_t)$.

Monotonicity

$$\tilde{\beta}_h = \int \tilde{\omega}_Z(x) g_h'(x)\, dx, \quad \tilde{\omega}_Z(x) \equiv \text{Cov}(\mathbb{1}\{X_t \geq x\}, \zeta(X_t)) / \text{Var}(Z_t)$$

- $\tilde{\omega}_Z(\cdot) \geq 0$ iff. "monotonicity on average":

$$E[Z_t \mid X_t \geq x] \geq E[Z_t \mid X_t < x] \quad \text{for all } x.$$

  - Implied by monotonicity of $\zeta(x) \equiv E[Z_t \mid X_t = x]$. Also implies $\tilde{\omega}_Z(\cdot)$ is hump-shaped.

- Lesson: proxies should be approximately monotonically related to shock of interest, but relationship need not be close to linear.
  - E.g., "narrative sign restriction": $Z_t = \mathbb{1}\{X_t \geq c_2\} - \mathbb{1}\{X_t \leq -c_1\}$. Antolín-D & Rubio-R (2018)

- In paper: our monotonicity condition is much weaker than "uniform monotonicity" cond'n required when X_t is endogenous. Imbens & Angrist (1994); Angrist, Graddy & Imbens (2000)

Outline

Identification via heteroskedasticity

- When shocks/proxies are not available, popular to identify latent shock X via restrictions on shock heterosk'y in linear SVAR. Sentana & Fiorentini (2001); Rigobon (2003); Lewis (2024)
- We consider nonparametric analogue of this approach to show sensitivity to linearity.
- Observe $(\mathbf{Y}, D)$ from **nonparametric factor model** (drop time subscripts):

$$\mathbf{Y} = \psi(X, \mathbf{U}), \quad (D, X) \perp\!\!\!\perp \mathbf{U}.$$

- D: regime. Proxy for X, but does not affect mean, only variance and higher moments:

$$E[X \mid D] = 0, \quad X \not\perp\!\!\!\perp D.$$

Identification via heteroskedasticity

- When shocks/proxies are not available, popular to identify latent shock X via restrictions on shock heterosk'y in linear SVAR. Sentana & Fiorentini (2001); Rigobon (2003); Lewis (2024)

- We consider nonparametric analogue of this approach to show sensitivity to linearity.

- Observe $(\mathbf{Y}, D)$ from **nonparametric factor model** (drop time subscripts):

$$\mathbf{Y} = \psi(X, \mathbf{U}), \quad (D, X) \perp\!\!\!\perp \mathbf{U}.$$

- D: regime. Proxy for X, but does not affect mean, only variance and higher moments:

$$E[X \mid D] = 0, \quad X \not\perp\!\!\!\perp D.$$

- Linear ID: If $\psi(x, \mathbf{u}) = \boldsymbol{\theta} x + \gamma(\mathbf{u})$ and $\theta_1 \operatorname{Cov}(X^2, D) \neq 0$, then

$$\frac{\operatorname{Cov}(\mathbf{Y}, Z)}{\operatorname{Cov}(Y_1, Z)} = \frac{\boldsymbol{\theta}}{\theta_1}, \quad \text{where} \quad Z \equiv (D - E[D])Y_1.$$

Rigobon & Sack (2004); Lewbel (2012)

ID via heteroskedasticity: large nonparametric identified set

- Proposition: Suppose we observe $(\mathbf{Y}, D)$ from the nonparametric model

$$\mathbf{Y} = \psi(X, \mathbf{U}), \quad X = \sigma(D)W, \quad W \perp\!\!\!\perp D \perp\!\!\!\perp \mathbf{U},$$

  where $\sigma(\cdot) \geq 0$ is known, the distr'n of W is symmetric around 0 and known, and the number $m \geq 2$ of shocks is known.

  Then the identified set for $\psi(x, \mathbf{u})$ contains a function that is symmetric around 0 in x.

- Can never rule out zero causal effect, $\int \omega(x) \frac{\partial E[\psi(x,\mathbf{U})]}{\partial x} \, dx = 0$, for symmetric ω!
- Intuition: D shifts only scale of X, not location $\implies$ can't construct proxy Z that satisfies monotonicity requirement, without imposing fct'l form as'ns on ψ.
  - Known issue in linear ID: shock variance depends on regime, yet require coef's to be constant. In nonparametric context, there's no distinction between "shock variances" and "coefficients".

ID via heteroskedasticity: sensitivity of linear procedures

- Proposition: Assume additively separable model

$$\mathbf{Y} = \boldsymbol{\theta}(X) + \gamma(\mathbf{U}).$$

Then Rigobon-Sack-Lewbel instrument $Z \equiv (D - E[D])Y_1$ satisfies

$$\text{Cov}(\mathbf{Y}, Z) = \int \check{\omega}(x)\boldsymbol{\theta}'(x)\,dx, \quad \text{for weights } \check{\omega}(x) \text{ that}\ldots$$

  - ...can integrate to 0 $\Longrightarrow$ estimate 0 causal effect of X on Y_j even if $\theta_j(x)$ is linear!
  - ...can be negative even in favorable case $Y_1 = X$, depending on entire distribution $(X \mid D)$.

- In non-separable models, we may not estimate *any* weighted avg of causal effects: if $\mathbf{Y} = X\gamma(\mathbf{U})$ with $E[\gamma(\mathbf{U})] = \mathbf{0}$, then $E[\mathbf{Y} \mid X] = \mathbf{0}$ but $\text{Cov}(\mathbf{Y}, Z) \neq 0$.

ID via heteroskedasticity: sensitivity of linear procedures

- Proposition: Assume additively separable model

$$\mathbf{Y} = \boldsymbol{\theta}(X) + \boldsymbol{\gamma}(\mathbf{U}).$$

Then Rigobon-Sack-Lewbel instrument $Z \equiv (D - E[D])Y_1$ satisfies

$$\text{Cov}(\mathbf{Y}, Z) = \int \check{\omega}(x)\boldsymbol{\theta}'(x)\, dx, \quad \text{for weights } \check{\omega}(x) \text{ that...}$$

  - ...can integrate to 0 $\implies$ estimate 0 causal effect of X on Y_j even if $\theta_j(x)$ is linear!
  - ...can be negative even in favorable case $Y_1 = X$, depending on entire distribution $(X \mid D)$.

- In non-separable models, we may not estimate *any* weighted avg of causal effects: if $\mathbf{Y} = X\boldsymbol{\gamma}(\mathbf{U})$ with $E[\boldsymbol{\gamma}(\mathbf{U})] = \mathbf{0}$, then $E[\mathbf{Y} \mid X] = \mathbf{0}$ but $\text{Cov}(\mathbf{Y}, Z) \neq 0$.
- Silver lining: at least linearity is testable. Power? Rigobon & Sack (2004); Wright (2012)

Outline

Identification via non-Gaussianity

- Recently popular procedure in linear SVAR literature: identify latent shocks by assuming they are independent and non-Gaussian. Gouriéroux, Monfort & Renne (2017); Lanne, Meitz & Saikkonen (2017); Lewbel, Schennach & Zhang (2024); Lewis (2024)
  - A.k.a. "independent components analysis" (ICA) outside economics. Comon (1994)
- Start from nonparametric factor model, but now we only observe $\mathbf{Y}$ (no proxy):

$$\mathbf{Y} = \psi(X, \mathbf{U}), \quad X \perp\!\!\!\perp \mathbf{U}.$$

Identification via non-Gaussianity

- Recently popular procedure in linear SVAR literature: identify latent shocks by assuming they are independent and non-Gaussian. Gouriéroux, Monfort & Renne (2017); Lanne, Meitz & Saikkonen (2017); Lewbel, Schennach & Zhang (2024); Lewis (2024)
  - A.k.a. "independent components analysis" (ICA) outside economics. Comon (1994)
- Start from nonparametric factor model, but now we only observe $\mathbf{Y}$ (no proxy):

$$\mathbf{Y} = \psi(X, \mathbf{U}), \quad X \perp\!\!\!\perp \mathbf{U}.$$

- Linear ID (Darmois-Skitovich theorem):
  - Assume $\psi(x, \mathbf{u}) = \boldsymbol{\theta} x + \boldsymbol{\gamma}\mathbf{u}$, and the shocks $(X, U_1, \dots, U_{m-1})$ are independent and non-Gaussian (except perhaps one).
  - Then two linear combinations $\varsigma'\mathbf{Y}$ and $\tilde{\varsigma}'\mathbf{Y}$ of the data can only be independent if they equal two different shocks (up to sign and scale).

ID via non-Gaussianity: large nonparametric identified set

$$\mathbf{Y} = \psi(X, \mathbf{U}), \quad X \perp\!\!\!\perp \mathbf{U}$$

- Unfortunately, there is no general nonlinear Darmois-Skitovich theorem: the nonparametric identified set for the above model is huge. Jutten & Karhunen (2003)
- Problem: independence and non-Gaussianity as'ns are vacuous in nonparametric context.
  - Can always transform a uniform r.v. into any distribution via the quantile function.
  - Can always transform one uniform r.v. into two independent uniforms.
  - In particular, we can represent $\mathbf{Y} = \tilde{\psi}(X)$ where $X \sim \text{unif}([0, 1]) \implies$ can't rule out that X drives all the variation in all observed variables!
  - Formal results in paper.

ID via non-Gaussianity: sensitivity of linear procedures

- Easy to construct cases where any linear ICA procedure is inconsistent *and* the linear model is unfalsifiable.

- Example: Suppose $(X, U) \sim N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$ and

$$Y_1 \equiv X + U, \quad Y_2 \equiv \gamma(X - U),$$

where $\gamma(\cdot)$ is an arbitrary nonlinear fct.

  - Interpretation: linear ICA model, but we got transformation of Y_2 slightly wrong.
  - $Y_1 \perp\!\!\!\perp Y_2 \implies$ linear ICA procedure concludes that $Y_1 =$ "shock 1" and $Y_2 =$ "shock 2". Nothing in the data can reject the linear model.
  - But X actually only contributes 50% of the variance of Y_1.
  - Discontinuity: same (asymptotic) bias regardless of how close $\gamma(\cdot)$ is to linear.

Outline

1. Nonparametric framework for dynamic causality
2. The Good: observed shocks and proxies
   - Observed shocks
   - Proxies
3. The Bad: identification via heteroskedasticity
4. The Ugly: identification via non-Gaussianity
5. Identification of average marginal effects
6. Conclusion

General result on ID of average marginal effects

- How can we identify avg marginal effects for pre-specified weight fct ω?
- Outcome Y, regressor X (arbitrary distribution!), covariates $\mathbf{W}$. Define

$$g(x, \mathbf{w}) \equiv E[Y \mid X = x, \mathbf{W} = \mathbf{w}].$$

  - If X has gaps in support (e.g., discrete/mixed), extend g to an interval via linear interpolation.
  - $g'(x, \mathbf{w})$: derivative wrt. x $(= g(1, \mathbf{w}) - g(0, \mathbf{w})$ for binary $X)$.
- Proposition: Under weak regularity conditions, for any α s.t. $E[\alpha(X, \mathbf{W}) \mid \mathbf{W}] = 0$,

$$E[\alpha(X, \mathbf{W})Y] = E[\alpha(X, \mathbf{W})g(X, \mathbf{W})] = E\left[\int \omega(x, \mathbf{W})g'(x, \mathbf{W})\, dx\right],$$

where $\omega(x, \mathbf{w}) \equiv E[\mathbb{1}\{X \geq x\}\alpha(X, \mathbf{W}) \mid \mathbf{W} = \mathbf{w}]$. Newey & Stoker (1993)

Identification of average marginal effects: implications

$$E[\alpha(X, \mathbf{W})Y] = E\left[\int \omega(x, \mathbf{W}) g'(x, \mathbf{W})\, dx\right] \qquad (\dagger)$$

- Implies most of the preceding propositions.
- With covariates, can derive representation of estimand from partially linear regression as weighted avg of marginal effects. (More in paper.)

$$Y = X\beta + \gamma(\mathbf{W}) + \text{residual}, \quad \text{where} \quad \gamma \in \Gamma \text{ (potentially nonparametric class).}$$

  - X can be discrete/cts/mixed. Special cases: Angrist & Krueger (1999); de Chaisemartin & D'Haultfœuille (2020); Goodman-Bacon (2021); Goldsmith-Pinkham, Hull & Kolesár (2024)
- For given ω, estimate average marginal effect on RHS of ($\dagger$) by reverse-engineering Riesz representer α and reporting the weighted outcome on the LHS of ($\dagger$).
  - α will generally require nonparametric estimation. Recent double-robust/debiased ML literature suggests combining weighting with outcome modeling. (More in paper.)

Outline

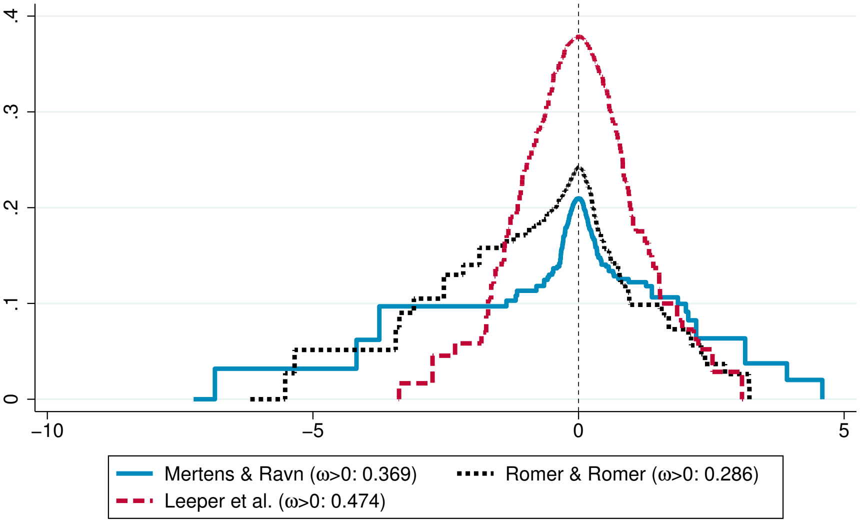
1. Nonparametric framework for dynamic causality
2. The Good: observed shocks and proxies
   - Observed shocks
   - Proxies
3. The Bad: identification via heteroskedasticity
4. The Ugly: identification via non-Gaussianity
5. Identification of average marginal effects
6. Conclusion

Conclusion

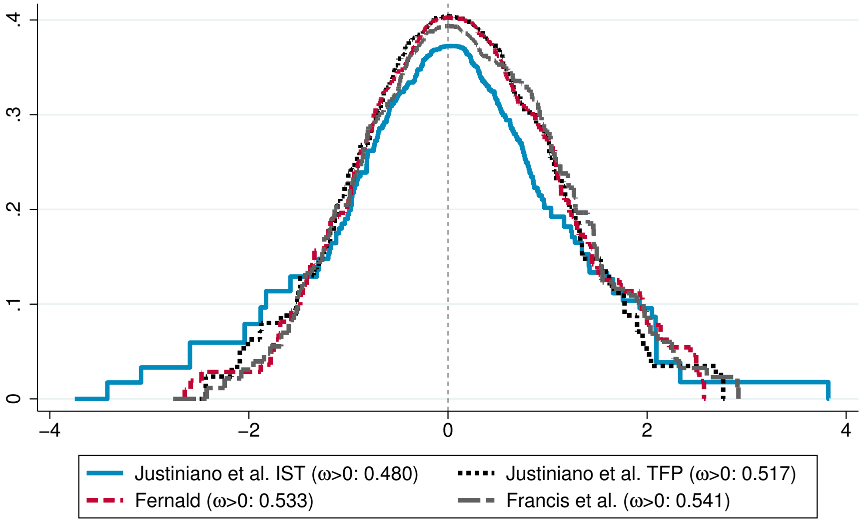
- Hard work of constructing shock measures/proxies pays off: robustness to nonlinearity.
  - Report implied causal weight function (Stata code in our GitHub repo).
  - Proxies should be approx'ly monotonically related to shocks, but not necessarily linearly.
  - If using covariates for identification, check sensitivity wrt. functional form.
- Identification approaches based on latent shocks sensitive to linearity assumption.
  - This paper: ID via heterosk'y/non-Gauss'y. Future work: ID via long-run/sign restrictions.
- Nonparametric TE literature has useful lessons for macro, despite our smaller data sets.
  Angrist & Kuersteiner (2011); Angrist, Jordà & Kuersteiner (2018); Rambachan & Shephard (2021)

Appendix

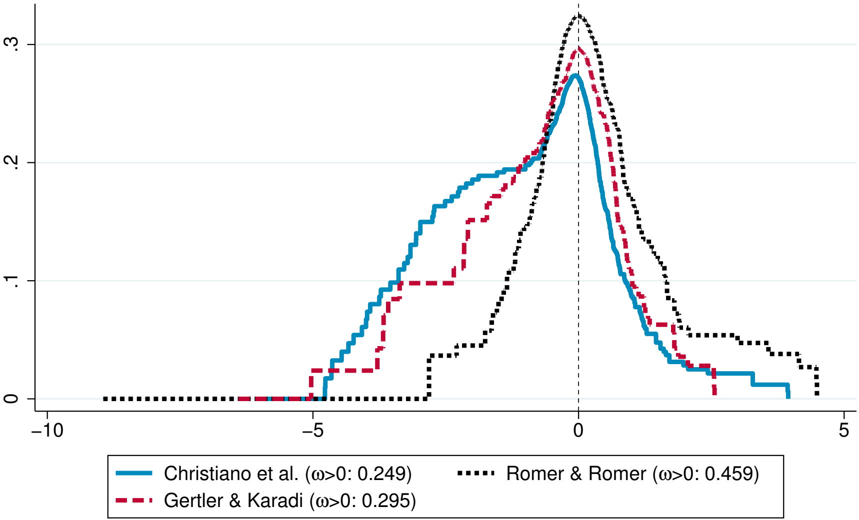
Causal weight functions: tax shocks



Causal weight functions: technology shocks



Causal weight functions: monetary policy shocks



ID via heteroskedasticity: testable restrictions

- While ID via heteroskedasticity is sensitive to linearity, at least linearity is testable.
- If $\mathbf{Y} = \boldsymbol{\theta} X + \gamma(\mathbf{U})$ and $(D, X) \perp\!\!\!\perp \mathbf{U}$, then

$$\text{Var}(\mathbf{Y} \mid D = d_1) - \text{Var}(\mathbf{Y} \mid D = d_0) = [\text{Var}(X \mid D = d_1) - \text{Var}(X \mid D = d_0)]\boldsymbol{\theta}\boldsymbol{\theta}'$$

  should be a rank-1 matrix. Rigobon & Sack (2004); Wright (2012)
- Power against nonlinear alternatives?

ID via non-Gaussianity: second example of sensitivity

- Example: Only nonlinearity is relationship btw Y_2 and U,

$$Y_1 = X + U, \quad Y_2 = X + \gamma(U), \quad X \perp\!\!\!\perp U.$$

- Can choose distr'ns for X and U and a nonlinear fct γ s.t. $Y_1 \perp\!\!\!\perp Y_2$.
- Then any linear ICA procedure erroneously concludes $Y_1 =$ "shock 1", $Y_2 =$ "shock 2".
- Proof: by Box-Muller transform, with $\tilde{U}_1$ and $\tilde{U}_2$ independent uniforms,

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

Hence, we can set

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

$$X \equiv \log(-2\log \tilde{U}_1), \quad U \equiv \log\cos^2(2\pi\tilde{U}_2), \quad \gamma(u) \equiv \log\left(1 - \exp(u)\right).$$