Different estimation strategies and nonlinearities

Tomás E. Caravello
MIT

EACBN, June 2024

Outline

Outline

1. A Menu of Estimation Strategies

2. Nonlinearities
    Non-Parametric Techniques

Running Example: monetary policy shocks

- **Variables:** GDP (y_t), Unemployment (u_t), Inflation (π_t), Nominal Rate (i_t).

- **Identification:** use Aruoba and Drechsel (2024) shock (z_t^m).
    - For VAR-based, add it to the VAR first and use recursive ID.
    - For LP-based, use it as perfect measure of the shock.
      [Equivalent to LP-IV if we normalize interest rate response]
    - In all cases, include a constant and 4 lags of all variables.

- We discuss several **estimation** strategies.
    - **VAR-based**: Standard VAR, Bias Corrected VAR, Bayesian VAR.
    - **LP-based**: LP, Bias Corrected LP, Penalized LP.

- **Goal:** provide a (non-extensive) summary of some recent methods
  Disclaimer: won't cover inference, large literature, can provide references

VAR-based: Standard and Bias-Correction

- Let $y_t = [z_t^m, y_t, u_t, \pi_t, i_t]$.
- **Standard VAR**
  - Fit by OLS
$$y_t = c + \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + u_t \tag{1}$$
    get $\hat{A}_\ell$, $\hat{\Sigma} = \widehat{Var}[u_t]$.
  - Use Choslesky Decomposition to get $\hat{C}\hat{C}' = \hat{\Sigma}$
  - Response at time 0 is first column of $\hat{C}$. Get other horizons by iterating on (1).
  - If desired, rescale IRFs to get unit response in some variable.

VAR-based: Standard and Bias-Correction

- Let $y_t = [z_t^m, y_t, u_t, \pi_t, i_t]$.
- **Standard VAR**
  - Fit by OLS
  $$y_t = c + \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + u_t \tag{1}$$
  get $\hat{A}_\ell$, $\hat{\Sigma} = \widehat{Var}[u_t]$.
  - Use Choslesky Decomposition to get $\hat{C}\hat{C}' = \hat{\Sigma}$
  - Response at time 0 is first column of $\hat{C}$. Get other horizons by iterating on (1).
  - If desired, rescale IRFs to get unit response in some variable.
- **Bias-Corrected VAR**
  - It turns out that OLS estimates $\hat{A}_\ell$ have T^{-1} order bias (Pope, 1990):
  $$E[\hat{A}_\ell - A] = -\mathbb{B}_\ell / T + O(T^{-3/2})$$
  where $\mathbb{B}_\ell$ can be computed and corrected. Then proceed as above.
  See Kilian and Lütkepohl (2017) ch 2.3.3 for details.

VAR-based: Bayesian VAR

- Treat (1) as a Normal, Inverse Wishart Bayesian regression.
- Assume $u_t \sim N(0, \Sigma)$, collect $\beta = vec([c, A_1, \ldots, A_p]')$.
- **Prior**

$$\begin{aligned} \Sigma &\sim IW(\Psi, d) \\ \beta|\Sigma &\sim N(b, \Sigma \otimes \Omega\lambda^2) \end{aligned}$$

VAR-based: Bayesian VAR

- Treat (1) as a Normal, Inverse Wishart Bayesian regression.
- Assume $u_t \sim N(0, \Sigma)$, collect $\beta = vec([c, A_1, \ldots, A_p]')$.
- **Posterior**:

$$
\begin{aligned}
\beta|\Sigma, y &\sim N(\hat{\beta}(\lambda^2), \hat{V}(\lambda^2)) \\
\hat{\beta}(\lambda) &= (x'x + (\Omega\lambda^2)^{-1})^{-1}(x'y + (\Omega\lambda^2)^{-1}\bar{b}) \\
\hat{V}(\lambda) &= \Sigma \otimes (x'x + (\Omega\lambda^2)^{-1})^{-1}
\end{aligned}
$$

where $y = [y_p, \ldots, y_T]$ and x contains a constant and p lags of y_t.

VAR-based: Bayesian VAR

- Treat (1) as a Normal, Inverse Wishart Bayesian regression.
- Assume $u_t \sim N(0, \Sigma)$, collect $\beta = vec([c, A_1, \ldots, A_p]')$.
- **Posterior**:

$$\begin{aligned} \beta|\Sigma, y &\sim N(\hat{\beta}(\lambda^2), \hat{V}(\lambda^2)) \\ \hat{\beta}(\lambda) &= (x'x + (\Omega\lambda^2)^{-1})^{-1}(x'y + (\Omega\lambda^2)^{-1}\bar{b}) \\ \hat{V}(\lambda) &= \Sigma \otimes (x'x + (\Omega\lambda^2)^{-1})^{-1} \end{aligned}$$

  where $y = [y_p, \ldots, y_T]$ and x contains a constant and p lags of y_t.
- High λ: loose prior, little shrinkage. [Note: $\lambda \to \infty$ yields OLS estimates]

VAR-based: Bayesian VAR

- Treat (1) as a Normal, Inverse Wishart Bayesian regression.
- Assume $u_t \sim N(0, \Sigma)$, collect $\beta = vec([c, A_1, \ldots, A_p]')$.
- **Posterior**:

$$
\begin{aligned}
\beta|\Sigma, y &\sim N(\hat{\beta}(\lambda^2), \hat{V}(\lambda^2)) \\
\hat{\beta}(\lambda) &= (x'x + (\Omega\lambda^2)^{-1})^{-1}(x'y + (\Omega\lambda^2)^{-1}\bar{b}) \\
\hat{V}(\lambda) &= \Sigma \otimes (x'x + (\Omega\lambda^2)^{-1})^{-1}
\end{aligned}
$$

  where $y = [y_p, \ldots, y_T]$ and x contains a constant and p lags of y_t.
- Low λ: tight prior, high shrinkage.

VAR-based: Bayesian VAR

- Treat (1) as a Normal, Inverse Wishart Bayesian regression.
- Assume $u_t \sim N(0, \Sigma)$, collect $\beta = vec([c, A_1, \ldots, A_p]')$.
- **Posterior**:

$$\begin{aligned}
\beta|\Sigma, y &\sim N(\hat{\beta}(\lambda^2), \hat{V}(\lambda^2)) \\
\hat{\beta}(\lambda) &= (x'x + (\Omega\lambda^2)^{-1})^{-1}(x'y + (\Omega\lambda^2)^{-1}\bar{b}) \\
\hat{V}(\lambda) &= \Sigma \otimes (x'x + (\Omega\lambda^2)^{-1})^{-1}
\end{aligned}$$

  where $y = [y_p, \ldots, y_T]$ and x contains a constant and p lags of y_t.
- How do we choose optimal λ?
- Giannone et al. (2015): derive the likelihood as function of hyperparameters. Then do Bayesian analysis.

LP-based: Standard and Bias-Correction

- **Standard LP**
  - Fit by OLS
$$y_{j,t+h} = c_{j,h} + \beta_{j,h} z_t^m + \text{ controls } + u_t \tag{2}$$
    for all horizons and variables of interest. Obtain $\{\hat{\beta}_{j,h}\}$
  - If desired, rescale IRFs to get unit response in some variable.

LP-based: Standard and Bias-Correction

- **Standard LP**
  - Fit by OLS
    $$y_{j,t+h} = c_{j,h} + \beta_{j,h} z_t^m + \text{ controls } + u_t \tag{2}$$
    for all horizons and variables of interest. Obtain $\{\hat{\beta}_{j,h}\}$
  - If desired, rescale IRFs to get unit response in some variable.
- **Bias-Corrected LP**
  - Herbst and Johannsen (2024) show that $\hat{\beta}_{j,h}$ also has T^{-1} bias
    $$E[\hat{\beta}_{j,h} - \beta_{j,h}] = -\mathbb{B}_{j,h}^{LP}/T + O(T^{-3/2})$$
    where $\mathbb{B}_{j,h}^{LP}$ can be computed and corrected. Then proceed as above.
  - Bias can be substantial if T is small.

LP-based: Penalized Local Projections

- LP is very noisy, since we estimate $\beta_{j,h}$ for each horizon independently.

LP-based: Penalized Local Projections

- Barnichon and Brownlees (2019): put some structure on the $\beta_{j,h}$ coefficients.

LP-based: Penalized Local Projections

- Barnichon and Brownlees (2019): put some structure on the $\beta_{j,h}$ coefficients.
- In particular, assume $\beta_{j,h} \approx \sum_{k=1}^{K} b_{j,k} B_k(h)$, where $B_k(x)$ is a set of B-spline basis.
  We'll see why they choose this in a second.

LP-based: Penalized Local Projections

- Barnichon and Brownlees (2019): put some structure on the $\beta_{j,h}$ coefficients.
- In particular, assume $\beta_{j,h} \approx \sum_{k=1}^{K} b_{j,k} B_k(h)$, where $B_k(x)$ is a set of B-spline basis.
- Ignore controls. Doing the same for the constant, we can write:

$$y_{j,t+h} \approx \sum_{k=1}^{K} c_{j,k} B_k(h) + \sum_{k=1}^{K} b_{j,k} B_k(h) z_t^m + u_t \tag{3}$$

LP-based: Penalized Local Projections

- Barnichon and Brownlees (2019): put some structure on the $\beta_{j,h}$ coefficients.
- In particular, assume $\beta_{j,h} \approx \sum_{k=1}^{K} b_{j,k} B_k(h)$, where $B_k(x)$ is a set of B-spline basis.
- Ignore controls. Doing the same for the constant, we can write:

$$y_{j,t+h} \approx \sum_{k=1}^{K} c_{j,k} B_k(h) + \sum_{k=1}^{K} b_{j,k} B_k(h) z_t^m + u_t \tag{3}$$

- Stack this for all h as:

$$\mathcal{Y}_t = \mathcal{X}_t \theta + \mathcal{U}_t \tag{4}$$

Where $\theta_j = (c_{j,1}, \cdots, c_{j,K}, b_{j,1}, \cdots, b_{j,K})$, $\mathcal{Y}_t = [y_t, \cdots, y_{t+H}]$ and $\mathcal{X}_t$ is a matrix that contains elements $B_k(k)$ and $B_k(h) z_t^m$

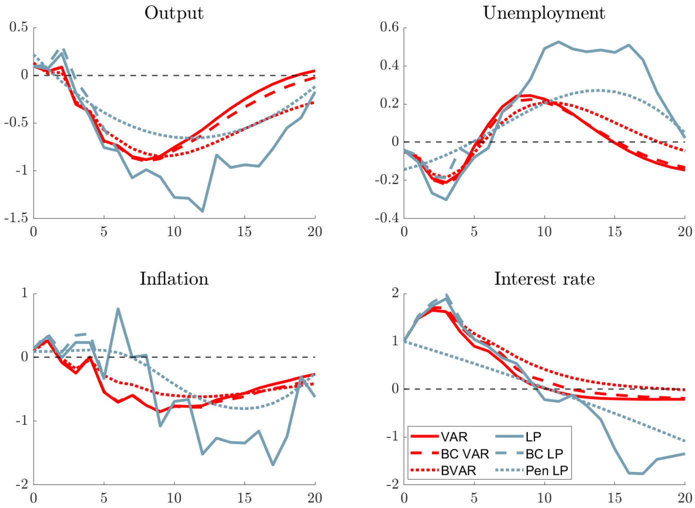
LP-based: Penalized Local Projections, cont.

- Penalized LP solves:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} ||\mathcal{Y} - \mathcal{X}\theta||^2 + \lambda\theta'\mathbf{P}\theta \tag{5}$$

- If $\lambda = 0$, this is OLS: we just rotated the regressors by writing them in other basis.
- With this basis, choosing $\mathbf{P}$ we can shrink the resulting β_h towards a polynomial.
- How do we choose λ? k-fold cross validation
- In our example, shrink towards quadratic.

IRFs to Aruoba-Drechsel shock



Discussion

- VARs and LP are only two out of a larger menu of estimation strategies.
- **Key trade-off**: bias and variance.
    - Imposing more structure (VARs, BVARs or Penalized LPs) reduces variance, but at the cost of bias.

Discussion

- VARs and LP are only two out of a larger menu of estimation strategies.
- **Key trade-off**: bias and variance.
  - Imposing more structure (VARs, BVARs or Penalized LPs) reduces variance, but at the cost of bias.
  - Li et al. (2024): In many cases, estimators other than pure VAR or LP are better.
    Pen LP retain conceptual simplicity of LPs but eliminate their noisy behavior.
    BVARs improve specially for short horizons, and can handle larger number of variables.

Discussion

- VARs and LP are only two out of a larger menu of estimation strategies.
- **Key trade-off**: bias and variance.
  - Imposing more structure (VARs, BVARs or Penalized LPs) reduces variance, but at the cost of bias.
  - Li et al. (2024): In many cases, estimators other than pure VAR or LP are better.
    Pen LP retain conceptual simplicity of LPs but eliminate their noisy behavior.
    BVARs improve specially for short horizons, and can handle larger number of variables.
  - Montiel Olea et al. (2024): however, the extra structure of VARs creates vulnerabilities.
    Under mispecification, confidence intervals are off-centered so coverage probability is incorrect.

Outline

1. A Menu of Estimation Strategies

2. Nonlinearities

   Non-Parametric Techniques

Parametric vs Non-parametric approaches

- **Parametric Approach:**
  1. Idea: commit to a parametric, non-linear (econometric) model of the macro-economy
     Often chosen to speak to a particular kind of possible non-linearity: state dependence, sign dependence, zero lower bound ▸ Examples
  2. Although popular, this approach has some issues:
     - **Relation to structural modeling**: linearized models imply SVMAs, but there are no similar guarantees for non-linear models. In fact globally solved DSGEs may not map into the typical reduced-form models in the literature. Model mis-specification is thus likely.
     - **Vulnerability to mis-specification**: identification & estimation rely heavily on functional form assumptions & the entire model being correctly specified. Not plausible in practice.
       Compare with our previous linear analysis—only second moments matter.
- Natural alternative: **non-parametric approaches**

Outline

1. A Menu of Estimation Strategies

2. Nonlinearities
   Non-Parametric Techniques

Recent push: bringing microeconometric insights to macro

- How should we think about potential outcomes in a time series context?
  Reference is Rambachan and Shephard (2021).
  - **Main takeaways**
    1. External instruments robustly estimate weighted average treatment effects. Internal instruments are more fragile.
       This point is quite general in econometrics, see nice summary by Hull (2024).
    2. What do standard (linear) LPs identify?
       Akin to cross-sectional literature on OLS: some weighted average treatment effects.
    3. What do **non-linear** LPs identify? How we can use them to learn about state, sign and size dependence?

Rambachan and Shephard (2021): General DGP

- **Notation**: scalar outcome y_t, (scalar) shock $\varepsilon_{1,t}$ of interest, $\varepsilon_{-1,t}$ vector of other shocks, $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon'_{-1,t}]'$. Generalize linear SVMA to:
$$y_t = g\left(\{\varepsilon_{t-\ell}\}_{\ell \geq 0}\right) \tag{6}$$
Solution of all macro models has this form.

Rambachan and Shephard (2021): General DGP

- **Notation**: scalar outcome y_t, (scalar) shock $\varepsilon_{1,t}$ of interest, $\varepsilon_{-1,t}$ vector of other shocks, $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon'_{-1,t}]'$. Generalize linear SVMA to:

$$y_t = g\left(\{\varepsilon_{t-\ell}\}_{\ell \geq 0}\right) \tag{6}$$

  Solution of all macro models has this form.
- Fix horizon of interest h, then:

$$y_{t+h} = g(\underbrace{\{\varepsilon_{t+i}\}_{i=1}^{h}}_{\text{shocks after } t}, \varepsilon_{1,t}, \underbrace{\varepsilon_{-1,t},}_{\text{other shocks at } t} \underbrace{\{\varepsilon_{t-j}\}_{i=0}^{\infty}}_{\text{lags}}) = g^h(\varepsilon_{1,t}, u_{h,t}) \tag{7}$$

  where $u_{h,t}$ contains everything other than $\varepsilon_{1,t}$.

Rambachan and Shephard (2021): General DGP

- **Notation**: scalar outcome y_t, (scalar) shock $\varepsilon_{1,t}$ of interest, $\varepsilon_{-1,t}$ vector of other shocks, $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon'_{-1,t}]'$. Generalize linear SVMA to:

$$y_t = g\left(\{\varepsilon_{t-\ell}\}_{\ell \geq 0}\right) \tag{6}$$

Solution of all macro models has this form.

- Fix horizon of interest h, then:

$$y_{t+h} = g(\underbrace{\{\varepsilon_{t+i}\}_{i=1}^{h}}_{\text{shocks after } t}, \varepsilon_{1,t}, \underbrace{\varepsilon_{-1,t},}_{\text{other shocks at } t} \underbrace{\{\varepsilon_{t-j}\}_{i=0}^{\infty}}_{\text{lags}}) = g^h(\varepsilon_{1,t}, u_{h,t}) \tag{7}$$

where $u_{h,t}$ contains everything other than $\varepsilon_{1,t}$.

- Define the *average structural function*

$$G_h(\varepsilon) = \mathbb{E}\left[y_{t+h} \mid \varepsilon_{1,t} = \varepsilon\right] = \mathbb{E}\left[g^h(\varepsilon_{1,t}, u_{h,t}) \mid \varepsilon_{1,t} = \varepsilon\right]$$

This function gives the counterfactual average value of y_{t+h} if the shock was equal to ε, averaging out over all other shocks.

Identification schemes in a non-paramtetric world

- Rambachan and Shephard (2021) essentially distinguish two main cases

  1. **Observe outcome & shock** [or IV]: the average structural function is then non-parametrically identified as

  $$G_h(\varepsilon) = \mathbb{E}\left[y_{t+h} \mid \varepsilon_{1,t} = \varepsilon\right]$$

     Could be estimated with enough (i.e., a huge amount of) data.

  2. **Only observe outcomes**: (heavily) restrict the functional form to make progress.
     [See Section 7.1: impose linearity + invertibility + X. Otherwise nonstandard interpretation.]

- We'll now dig deeper into 1.
  - Probably hopeless to fully characterize $G_h(\varepsilon)$, in particular if want to additionally condition on something else, e.g., state of the economy
  - Instead: can we give an interpretation to the estimand of *simple* estimation strategies?

Linear methods in non-linear models

- Suppose we observe $\varepsilon_{1,t}$ and run the linear local projection

$$y_{t+h} = \beta_h \varepsilon_{1,t} + \text{controls} + \text{error}$$

- Then what do we estimate for an arbitrary general non-linear DGP?

Linear methods in non-linear models

- Suppose we observe $\varepsilon_{1,t}$ and run the linear local projection

$$y_{t+h} = \beta_h \varepsilon_{1,t} + \text{controls} + \text{error}$$

- Then what do we estimate for an arbitrary general non-linear DGP?

  1. You in general estimate a **best linear approximation** to the underlying $g(\bullet)$ function
     Plagborg-Møller and Wolf (2021)

     - Formally: estimand are entries of a pseudo-SVMA $\sum_{\ell=0}^{\infty} \Theta_\ell^* \varepsilon_{t-\ell}$ where

$$(\Theta_0^*, \Theta_1^*, \dots) \in \operatorname*{argmin}_{\tilde{\Theta}_0, \tilde{\Theta}_1, \dots} \mathbb{E}\left[\left(g(\{\varepsilon_{t-\ell}\}_{\ell \geq 0}) - \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \varepsilon_{t-\ell}\right)^2\right]$$

     - Analogous to best linear predictor property of OLS. Can we say more?

Linear methods in non-linear models

- Suppose we observe $\varepsilon_{1,t}$ and run the linear local projection

$$y_{t+h} = \beta_h \varepsilon_{1,t} + \text{controls} + \text{error}$$

- Then what do we estimate for an arbitrary general non-linear DGP?

  2. You in general estimate a particular **weighted average of causal effects for small shocks**

     - Again let $G_h(\epsilon)$ denote the average counterfactual future outcome for y if the shock—e.g., a policy instrument—was set to some value ϵ: $G_h(\epsilon) \equiv \mathbb{E}[y_{t+h} \mid \varepsilon_{1,t} = \epsilon]$
     - Then, as we will show on the next slide:

$$\beta_h = \frac{\text{Cov}(y_{t+h}, \varepsilon_t)}{\text{Var}(\varepsilon_t)} = \lim_{\delta \to 0} \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega(\epsilon) \frac{G_h(\epsilon + \delta) - G_h(\epsilon)}{\delta} d\epsilon$$

       where $\omega(\epsilon)$ is a weight function (non-negative and integrates to one), $supp(\varepsilon_t) \subset (-\bar{\varepsilon}, \bar{\varepsilon})$
     - Practical takeaway: plot $\omega(\epsilon)$. Most weight on positive/negative/small/large shocks?

Linear methods in non-linear models

- Let $F(\bullet)$ denote the cdf of $\varepsilon_{1,t}$, assume mean zero. Begin with the numerator:

$$\begin{aligned}
\mathrm{Cov}(y_{t+h}, \varepsilon_t) &= \mathbb{E}\left[\mathbb{E}\left(y_{t+h} \mid \varepsilon_{1,t}\right)\varepsilon_{1,t}\right] = \mathbb{E}\left[G_h(\varepsilon_{1,t})\varepsilon_{1,t}\right] \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} G_h(\epsilon)\epsilon dF(\epsilon) = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \left(\int_{-\bar{\varepsilon}}^{\epsilon} G_h'(a)da\right)\epsilon dF(\epsilon) \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \underbrace{\left(\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \mathbf{1}_{a\leq\epsilon}\epsilon dF(\epsilon)\right)}_{=\mathrm{Cov}(1_{a\leq\epsilon},\epsilon)} G_h'(a)da
\end{aligned}$$

Linear methods in non-linear models

- Let $F(\bullet)$ denote the cdf of $\varepsilon_{1,t}$, assume mean zero. Begin with the numerator:

$$\begin{aligned}
\text{Cov}(y_{t+h}, \varepsilon_t) &= \mathbb{E}\left[\mathbb{E}\left(y_{t+h} \mid \varepsilon_{1,t}\right)\varepsilon_{1,t}\right] = \mathbb{E}\left[G_h(\varepsilon_{1,t})\varepsilon_{1,t}\right] \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} G_h(\epsilon)\epsilon dF(\epsilon) = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \left(\int_{-\bar{\varepsilon}}^{\epsilon} G_h'(a)da\right)\epsilon dF(\epsilon) \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \underbrace{\left(\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \mathbf{1}_{a\leq\epsilon}\epsilon dF(\epsilon)\right)}_{=\text{Cov}(1_{a\leq\epsilon},\epsilon)} G_h'(a)da
\end{aligned}$$

- Next consider the denominator:

$$\text{Var}(\varepsilon_{1,t}) = \text{Cov}(\varepsilon_{1,t}, \varepsilon_{1,t}) = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \text{Cov}(\mathbf{1}_{a\leq\epsilon}, \epsilon)da$$

- The overall weight function is thus

$$\omega(a) \equiv \frac{\text{Cov}(\mathbf{1}_{a\leq\epsilon}, \epsilon)}{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \text{Cov}(\mathbf{1}_{b\leq\epsilon}, \epsilon)db}$$

Note: if $\varepsilon_{1,t}$ is mean-0 and normal then it can be shown that $\omega(a) \propto f(a)$.

The arguments in this slide come from Yitzhaki (1996) and Plagborg-Møller and Kolesar (2024).

State dependence in non-linear models

- Now suppose we observe $\varepsilon_{1,t}$ and run the simple state-dependent local projection

$$y_{t+h} = \beta_{0,h}(1 - s_{t-1})\varepsilon_{1,t} + \beta_{1,h} s_{t-1}\varepsilon_{1,t} + \text{controls} + \text{error}$$

  - Interpretation: s_{t-1} here would typically be the state of the economy, e.g. recession versus expansion. E.g. see Ramey and Zubairy (2018).

- Then what do we estimate for an arbitrary general non-linear DGP?
  - Assume that $\varepsilon_{1,t}$ is also independent of s_{t-1}. Then by a similar argument to above:

$$\beta_{s,h} = \lim_{\delta \to 0} \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \tilde{\omega}(\epsilon, s) \frac{\tilde{G}_h(\epsilon + \delta, s) - \tilde{G}_h(\epsilon, s)}{\delta} d\epsilon$$

    where $\tilde{G}_h(\epsilon, s) \equiv \mathbb{E}[y_{t+h} \mid \varepsilon_{1,t} = \epsilon, s_{t-1} = s]$ and $\tilde{\omega}(\epsilon, s)$ now additionally conditions on s
  - Takeaway: we now estimate a weighted average of state-dependent causal effects

Sign and Size dependence in non-linear models

- Consider the (population) regressions:
$$y_{t+h} = \beta_{1,h}\varepsilon_t + \beta_{2,h} f\left(\varepsilon_t\right) + \text{controls} + \text{error}$$
Where $f(x)$ is some non-linear transformation, such as x^2, x^3 or $|x|$. (Ascari and Haber (2021),Ben Zeev, Ramey, and Zubairy (2023), Forni et al. (2023))

Sign and Size dependence in non-linear models

- Consider the (population) regressions:

$$y_{t+h} = \beta_{1,h}\varepsilon_t + \beta_{2,h} f(\varepsilon_t) + \text{controls} + \text{error}$$

- Caravello and Martinez-Bruera (2024): if the shock $\varepsilon_{1,t}$ has symmetric distribution, then we can test for sign and size separately by picking f appropriately.

Sign and Size dependence in non-linear models

- Consider the (population) regressions:

$$y_{t+h} = \beta_{1,h}\varepsilon_t + \beta_{2,h} f\left(\varepsilon_t\right) + \text{controls} + \text{error}$$

- Caravello and Martinez-Bruera (2024): if the shock $\varepsilon_{1,t}$ has symmetric distribution, then we can test for sign and size separately by picking f appropriately.
- Following a similar procedure as before:

Lemma 1 (Weighted average representation)

$$\beta_{n,h} = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_n(a) G_h'(a) da \quad n = 1, 2$$

where $\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_1(a) da = 1$ and $\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2(a) da = 0$

Even-Odd decomposition

- Decompose $m(a) = G_h'(a)$ as

$$m(a) = m^{Even}(a) + m^{Odd}(a)$$

Even-Odd decomposition

- Decompose $m(a) = G_h'(a)$ as

$$m(a) = m^{Even}(a) + m^{Odd}(a)$$

- Recall:
  - By definition, $m^{Even}(a) = m^{Even}(-a)$ and $m^{Odd}(a) = -m^{Odd}(-a)$

Even-Odd decomposition

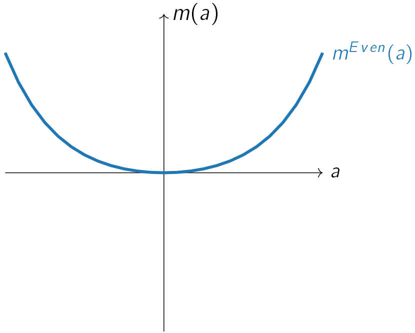
- Decompose $m(a) = G_h'(a)$ as

$$m(a) = m^{Even}(a) + m^{Odd}(a)$$

- Recall:
  - By definition, $m^{Even}(a) = m^{Even}(-a)$ and $m^{Odd}(a) = -m^{Odd}(-a)$
  - Decomposition always exists and is unique, given by $m^{Even}(a) = \frac{m(a)+m(-a)}{2}$ and $m^{Odd}(a) = \frac{m(a)-m(-a)}{2}$
  - Examples: x^2 and $|x|$ are even, x^3 is odd.

Even part captures only pure size differences

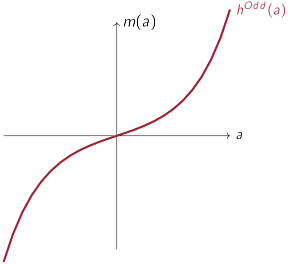
- Decompose $m(a) = G_h'(a)$ as $m(a) = m^{Even}(a) + m^{Odd}(a)$



- Since $m^{Even}(a) = m^{Even}(-a)$, **sign doesn't matter**

On average, odd part captures only sign differences

- Decompose $m(a) = G_h'(a)$ as $m(a) = m^{Even}(a) + m^{Odd}(a)$



- Since $m^{Odd}(a) = -m^{Odd}(-a) \Rightarrow \frac{m^{Odd}(a)+m^{Odd}(-a)}{2} = 0$
- **When averaging, size doesn't matter**
- Our estimands are averages, so this works well for our purposes!

Main Result

- We can also decompose $\omega_2(a) = \omega_2^{Odd}(a) + \omega_2^{Even}(a)$. Thus:

$$\beta_2 = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2(a) m(a) da$$

Main Result

- We can also decompose $\omega_2(a) = \omega_2^{Odd}(a) + \omega_2^{Even}(a)$. Thus:

$$\beta_2 = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2(a) m(a) da = \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Odd}(a) m^{Odd}(a) da}_{\text{Sign}} + \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Even}(a) m^{Even}(a) da}_{\text{Size}}$$

Main Result

- We can also decompose $\omega_2(a) = \omega_2^{Odd}(a) + \omega_2^{Even}(a)$. Thus:

$$\beta_2 = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2(a) m(a) da = \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Odd}(a) m^{Odd}(a) da}_{\text{Sign}} + \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Even}(a) m^{Even}(a) da}_{\text{Size}}$$

  - If $\omega_2(a)$ purely odd $\rightarrow$ only sign non-linearities!
  - If $\omega_2(a)$ purely even $\rightarrow$ only size non-linearities!

Main Result

- We can also decompose $\omega_2(a) = \omega_2^{Odd}(a) + \omega_2^{Even}(a)$. Thus:

$$\beta_2 = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2(a)m(a)da = \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Odd}(a)m^{Odd}(a)da}_{\text{Sign}} + \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Even}(a)m^{Even}(a)da}_{\text{Size}}$$

  - If $\omega_2(a)$ purely odd $\rightarrow$ only sign non-linearities!
  - If $\omega_2(a)$ purely even $\rightarrow$ only size non-linearities!
- Idea: pick the correct $f(\cdot)$

Main Result

- We can also decompose $\omega_2(a) = \omega_2^{Odd}(a) + \omega_2^{Even}(a)$. Thus:

$$\beta_2 = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2(a) m(a) da = \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Odd}(a) m^{Odd}(a) da}_{\text{Sign}} + \underbrace{\int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \omega_2^{Even}(a) m^{Even}(a) da}_{\text{Size}}$$

  - If $\omega_2(a)$ purely odd $\rightarrow$ only sign non-linearities!
  - If $\omega_2(a)$ purely even $\rightarrow$ only size non-linearities!
- Idea: pick the correct $f(\cdot)$

Proposition 1 Proof

Assume the unconditional distribution of ε_t is symmetric. Then

- If $f(\cdot)$ is even $\implies$ $\omega_2(a)$ is odd (i.e $\omega_2^{Even}(a) = 0 \quad \forall a$)
- If $f(\cdot)$ is odd $\implies$ $\omega_2(a)$ is even (i.e $\omega_2^{Odd}(a) = 0 \quad \forall a$)

Practical Takeaways

1. Sign non-linearities $\rightarrow$ use even transformation.
2. Size non-linearities $\rightarrow$ use odd transformation.
3. Test for $\beta_2 = 0$ using standard inference.
4. For interpreation, compute and plot weights (Plagborg-Møller and Kolesar, (2024)), and its even-odd decomposition.

Empirical Application: Oil Supply News Shocks

- We fit local projections using oil shock series (Känzig (2021)). Details

$$y_{t+h} = \alpha_h + \beta_{h,1}\varepsilon_t^{oil} + \beta_{h,2} f(\varepsilon_t^{oil}) + \text{ controls } + u_{t+h} \tag{8}$$

Empirical Application: Oil Supply News Shocks

- We fit local projections using oil shock series (Känzig (2021)). Details

$$y_{t+h} = \alpha_h + \beta_{h,1}\varepsilon_t^{oil} \qquad + \text{ controls } + u_{t+h} \tag{8}$$

- Linear: no $f(\cdot)$ term.

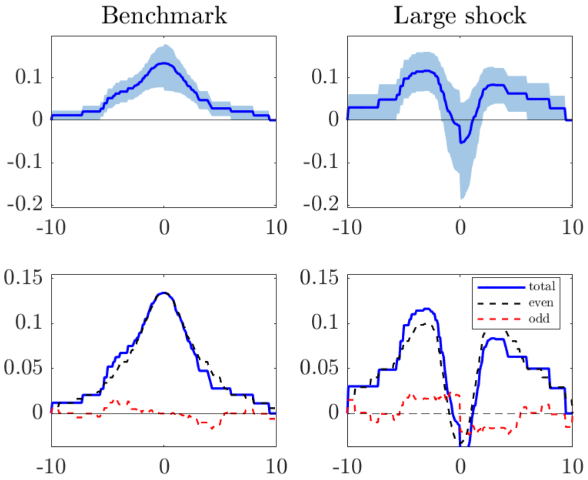
Empirical Application: Oil Supply News Shocks

- We fit local projections using oil shock series (Känzig (2021)).

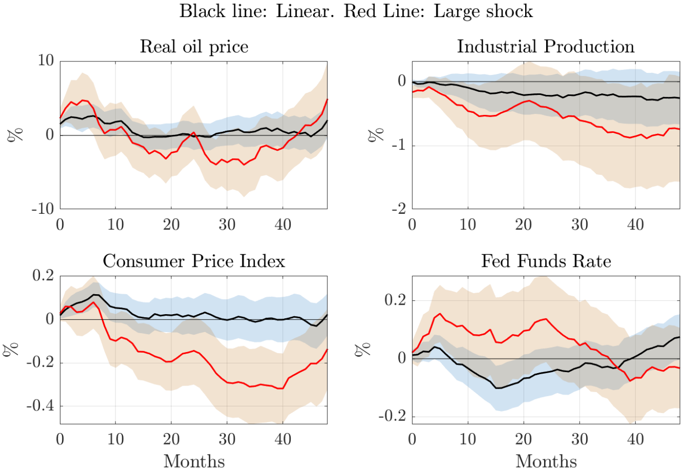
$$y_{t+k} = \alpha_k + \beta_{k,1}\varepsilon_t^{oil} + \beta_{k,2} f(\varepsilon_t^{oil}) + \text{ controls } + u_{t+k} \tag{9}$$

- For non-linear terms, use piecewise linear functions for ease of interpretation.
  - Sign: use $f(x) = |x|$
  - Size: use $f(x) = \mathbf{1}_{\{x \leq -\bar{b}\}}(x + \bar{b}) + \mathbf{1}_{\{x \geq \bar{b}\}}(x - \bar{b})$

Weights: linear only vs large shock ($\beta_{k,1} + \beta_{k,2}$)



Linear IRFs vs Large shock IRF ($\beta_{k,1} + \beta_{k,2}$)



Open issues

Literature on non-linear time series methods has recently seen a bit of a revival, still lots of interesting **open questions**

- **Methodological**
  - What kind of parametric models—if any—capture well the non-linearities implied by non-linearly solved DSGE models? [See Aruoba et al. (2017)]
  - How feasible are non-parametric estimators of general potential outcomes functions? [See Gonçalves et al. (2024)]
- **Applied**
  - Do non-linearities matter in the aggregate? If so, what are the relevant non-linearities? ZLB, downward wage rigidity, state-dependence in price setting or investment/durables?
  - What kind of non-linearities are most important? By size, sign, or state of the economy? E.g. Barnichon et al. (2022) for government spending multipliers.

References I

S. Borağan Aruoba and Thomas Drechsel. Identifying monetary policy shocks: A natural language approach. *University of Maryland, CEPR*, 2024. URL `http://econweb.umd.edu/~drechsel/papers/Aruoba_Drechsel.pdf`.

S. Borağan Aruoba, Luigi Bocola, and Frank Schorfheide. Assessing dsge model nonlinearities. *Journal of Economic Dynamics and Control*, 83:34–54, 2017. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2017.07.006. URL `https://www.sciencedirect.com/science/article/pii/S0165188917301562`.

Guido Ascari and Timo Haber. Non-Linearities, State-Dependent Prices and the Transmission Mechanism of Monetary Policy. *The Economic Journal*, 132(641):37–57, 06 2021. ISSN 0013-0133. doi: 10.1093/ej/ueab049. URL `https://doi.org/10.1093/ej/ueab049`.

References II

Regis Barnichon and Christian Brownlees. Impulse response estimation by smooth local projections. *The Review of Economics and Statistics*, 101(3):522–530, 2019. doi: 10.1162/rest_a_00843. URL https://www3.nd.edu/~nmark/Climate/BarnichonBrownless_ReStat2019.pdf.

Regis Barnichon, Davide Debortoli, and Christian Matthes. Understanding the size of the government spending multiplier: It's in the sign. *The Review of Economic Studies*, 89 (1):87–117, 2022. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdab029.

Nadav Ben Zeev, Valerie A. Ramey, and Sarah Zubairy. Do government spending multipliers depend on the sign of the shock? *AEA Papers and Proceedings*, 113:382–87, May 2023. doi: 10.1257/pandp.20231071. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20231071.

Tomás Caravello and Pedro Martinez-Bruera. Disentangling sign and size non-linearities. *SSRN*, January 2024. URL https://ssrn.com/abstract=4704050.

References III

Timothy Cogley and Thomas Sargent. Evolving post-world war ii u.s. inflation dynamics. *NBER Macroeconomics Annual*, 16:331–388, 2002. URL https://www.nber.org/system/files/chapters/c11068/c11068.pdf.

Mario Forni, Luca Gambettu, Nicolò Maffei-Faccioli, and Luca Sala. Nonlinear transmission of financial shocks: Some new evidence. *Journal of Money, Credit and Banking*, forthcoming, 2023. doi: https://doi.org/10.1111/jmcb.13099. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.13099.

Domenico Giannone, Michele Lenza, and Giorgio E. Primiceri. Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics*, 97(2):436–451, 05 2015. ISSN 0034-6535. doi: 10.1162/REST_a_00483. URL https://doi.org/10.1162/REST_a_00483.

References IV

Sílvia Gonçalves, Ana María Herrera, Lutz Kilian, and Elena Pesavento. State-dependent local projections. *Journal of Econometrics*, page 105702, 2024. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2024.105702. URL `https://www.sciencedirect.com/science/article/pii/S0304407624000484`.

Luca Guerrieri and Matteo Iacoviello. Occbin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics*, 70:22–38, 2015. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2014.08.005. URL `https://www.sciencedirect.com/science/article/pii/S0304393214001238`.

Edward P. Herbst and Benjamin K. Johannsen. Bias in local projections. *Journal of Econometrics*, 240(1):105655, 2024. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2024.105655. URL `https://www.sciencedirect.com/science/article/pii/S0304407624000010`.

References V

Lutz Kilian and Helmut Lütkepohl. *Structural Vector Autoregressive Analysis*. Cambridge University Press, 2017. ISBN 9781108164818. URL https://www.cambridge.org/core/books/structural-vector-autoregressive-analysis/DAF4217439EA585D10902D58A8849E06.

Diego R. Känzig. The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, 111(4):1092–1125, 2021. ISSN 0002-8282. doi: 10.1257/aer.20190964.

Dake Li, Mikkel Plagborg-Møller, and Christian K. Wolf. Local projections vs. vars: Lessons from thousands of dgps. *Journal of Econometrics*, page 105722, 2024. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2024.105722. URL https://www.sciencedirect.com/science/article/pii/S030440762400068X.

References VI

José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf. Double robustness of local projections and some unpleasant varithmetic. *arXiv preprint arXiv:2405.09509*, 2024. URL `https://arxiv.org/abs/2405.09509`.

Mikkel Plagborg-Møller and Christian K. Wolf. Local projections and VARs estimate the same impulse responses. *Econometrica*, 89(2):955–980, 2021. ISSN 1468-0262. doi: 10.3982/ECTA17813. URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17813`.

Alun Lloyd Pope. Biases of estimators in multivariate non-gaussian autoregressions. *Journal of Time Series Analysis*, 11(3):249–258, 1990. doi: 10.1111/j.1467-9892.1990.tb00056.x. URL `https://doi.org/10.1111/j.1467-9892.1990.tb00056.x`.

Giorgio E. Primiceri. Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852, 07 2005. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2005.00353.x. URL https://doi.org/10.1111/j.1467-937X.2005.00353.x.

Ashesh Rambachan and Neil Shephard. When do common time series estimands have nonparametric causal meaning? *working paper*, 2021.

Valerie A. Ramey and Sarah Zubairy. Government spending multipliers in good times and in bad: Evidence from us historical data. *Journal of Political Economy*, 126(2):850–901, 2018. doi: 10.1086/696277.

Shlomo Yitzhaki. On using linear regressions in welfare economics. *Journal of Business and Economic Statistics*, 14(4):478–486, 1996. ISSN 07350015. URL http://www.jstor.org/stable/1392256.

Appendix

Parametric Approach: Examples

Back

- We will briefly review a couple of the most notable examples:

1. **Regime-switching models**: model coefficients differ by *regime* of the macro-economy
2. **Time-varying parameter models**: model coefficients evolve continuously over time
3. **Censored autoregression** [mostly for ZLB constraint]

Regime-switching models

- **Framework**: allow parameters of model to differ by **regime**. E.g.:

$$y_t = \{A_0(1 - s_{t-1}) + A_1 s_{t-1}\} y_{t-1} + \eta_t, \quad \eta_t | s_{t-1} \sim N(0, \Sigma_0(1 - s_{t-1}) + \Sigma_1 s_{t-1})$$

- Popular special cases:

1. **Threshold Autoregression:**

$$s_t = \mathbf{1}_{x_t \leq c}$$

where x_t could be something exogenous or itself depend on y_t, and c is a parameter

2. **Smooth Transition Autoregression:**

$$s_t = \frac{1}{1 + \exp(\beta x_t - c)} \in [0, 1]$$

where again x_t may evolve exogenously or be a function of y_t, and β, c are parameters

3. **Markov Switching:** s_t evolves exogenously as a Markov Process.

Models with time-varying parameters

- **Framework**: parameters are **varying continuously**. E.g.:

$$y_t = A_t y_{t-1} + \Sigma_t^{1/2} \varepsilon_t, \quad \varepsilon_t \sim N(0, I_n)$$

- Popular special cases:
  - **Autoregressive component** $\beta_t \equiv \mathrm{vec}(A_t)$: independent random walks
  - **Volatility** Σ_t: multivariate stochastic volatility model
- Many well-known examples of these kind of approaches in the VAR literature
  Cogley and Sargent (2002),Primiceri (2005).

Models with censoring

- **Framework**: censoring of a subset of variables. E.g.:

$$\begin{aligned} y_{1,t} &= \max\{A_{1,\bullet} y_{t-1} + H_{1,\bullet} \varepsilon_t, 0\} \\ y_{2:n,t} &= A_{2:n,\bullet} y_{t-1} + H_{2:n,\bullet} \varepsilon_t \\ \varepsilon_t &\sim N(0, I_n) \end{aligned}$$

- Main application: binding **lower bound on nominal interest rates**
    - Assumption: linear system is adequate up to constraints on monetary policy. Echoes solution methods for structural models with occasionally binding constraints. See Guerrieri and Iacoviello (2015)

Summarizing $g(\bullet)$: dynamic causal effects

What are **dynamic causal effects** in these environments?

- No unique def'n. Note that such models generally imply recursive representations:

$$y_{t+h} = g_h\left(\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, \varepsilon_t, y_{t-1}\right)$$

- Suggests two possible natural defn's:

1. **Set other shocks to zero**

$$\theta_h(\delta_1, \delta_2; y_{t-1}) \equiv g_h(0, 0, \ldots, 0, \delta_2 \times e_1; y_{t-1}) - g_h(0, 0, \ldots, 0, \delta_1 \times e_1; y_{t-1})$$

2. **Average over other shocks**

$$\begin{aligned}\theta_h(\delta_1, \delta_2; y_{t-1}) \equiv \mathbb{E}_{\varepsilon_{t+h},\ldots,\varepsilon_{t+1},\varepsilon_{2:n,t}} [&g_h(\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, (\delta_2, \varepsilon_{2:n,t}); y_{t-1}) \\ &- g_h(\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, (\delta_1, \varepsilon_{2:n,t}); y_{t-1})]\end{aligned}$$

Note: definitions coincide in linear models and there depend only on $\delta_2 - \delta_1$.