Local Projections vs. VARs: Lessons From Thousands of DGPs

Dake Li  Mikkel Plagborg-Møller  Christian K. Wolf
Princeton University  Princeton University  MIT

Slides: https://scholar.princeton.edu/mikkelpm

October 16, 2021
Estimation of IRFs

- How to estimate **impulse response functions (IRFs)** in finite samples?

\[ \theta_h \equiv E(y_{t+h} \mid \varepsilon_{j,t} = 1) - E(y_{t+h} \mid \varepsilon_{j,t} = 0), \quad h = 0, 1, 2, \ldots \]

1. **Structural Vector Autoregression (VAR):** Sims (1980)

\[ w_t = \sum_{\ell=1}^{p} A_{\ell} w_{t-\ell} + B \varepsilon_t, \quad \varepsilon_t \sim WN(0, I_n). \]

Extrapolates \( \theta_h \) from first \( p \) autocovariances. Low variance, potentially high bias.

2. **Local Projections (LP):** Jordà (2005)

\[ y_{t+h} = \beta_h \varepsilon_{j,t} + \text{controls} + \text{residual}_{h,t}, \quad h = 0, 1, 2, \ldots \]

Estimates \( \theta_h \) from sample autocovariances out to lag \( h \). Low bias, high variance.
Choice of LP or VAR seems to matter for important applied questions. \textit{Ramey (2016)}

LP and VAR share same population IRF estimand at horizons $h \leq p$ (lag length). \textit{Plagborg-Møller & Wolf (2021)}

- No meaningful trade-off if interest centers on short horizons . . .
- . . .or if we choose very large lag length (high variance).


Analytical guidance is murky: Under local misspecification of VAR($p$) model, bias-variance trade-off depends on numerous aspects of DGP. \textit{Schorfheide (2005)}
This paper

• Our approach: **Large-scale simulation study** of impulse response estimators.

  • Draw 1,000s of DGPs from empirical Dynamic Factor Model. *Stock & Watson (2016)*

  • Several estimation methods: LP, VAR, shrinkage variants, . . .

  • Several identification schemes: observed shock, recursive, proxy/instrument.

  • Pay attention to researcher’s loss function and role of horizon.

• Question: Which estimators perform well on average across many DGPs?
1 Analytical example

2 Data generating processes

3 Estimators

4 Results

5 Conclusion
Simple analytical example

- Locally misspecified VAR(1) in the data $w_t \equiv (\varepsilon_{1,t}, y_t)'$:

$$y_t = \rho y_{t-1} + \varepsilon_{1,t} + \varepsilon_{2,t} + \frac{\alpha}{\sqrt{T}} \varepsilon_{2,t-1}, \quad (\varepsilon_{1,t}, \varepsilon_{2,t})' \sim N(0, \text{diag}(1, \sigma_2^2)).$$

- Parameter of interest: $\theta_h \equiv \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}} = \rho^h$.

- Two estimators (later consider other ones):

1. LP: $y_{t+h} = \hat{\beta}_h \varepsilon_{1,t} + \hat{\zeta}_h' w_{t-1} + \text{residual}_{h,t}$.

2. VAR: $w_t = \hat{A} w_{t-1} + \hat{C} \hat{\eta}_t$, where $\hat{C} = \text{Cholesky}$. Impulse response estimate $\hat{\delta}_h \propto e_2 \hat{A}^h \hat{C} e_1$ normalized so first variable $w_{1,t}$ responds by 1 unit on impact.
LP has zero asymptotic bias because it projects $y_{t+h}$ directly on shock $\varepsilon_{1,t}$.

VAR extrapolates, which lowers variance at the cost of bias when $\alpha \neq 0$. 

Asymptotic bias and standard deviation (Schorfheide, 2005)
How much should we care about bias to pick LP over VAR?

- Given loss function

$$\mathcal{L}_\omega(\theta_h, \hat{\theta}_h) = \omega \times \left( \mathbb{E}[\hat{\theta}_h - \theta_h] \right)^2 + (1 - \omega) \times \text{Var}(\hat{\theta}_h),$$

how much weight $\omega = \omega^*_h$ should we attach to bias$^2$ to be indifferent btw. LP and VAR?
Even in simple DGP, bias-variance trade-off is non-trivial. Depends on...

- persistence $\rho$ and degree $\alpha$ of misspecification.
- bias weight $\omega$ in loss function.
- impulse response horizon $h$.

Our approach going forward:

- Study trade-off through simulations in thousands of empirically calibrated DGPs. Will inform us about empirically relevant “$\rho$” and “$\alpha$”.
- Enrich menu of estimation procedures to trace out bias-variance possibility frontier.
Outline

1. Analytical example
2. Data generating processes
3. Estimators
4. Results
5. Conclusion
Encompassing model

• Empirical Dynamic Factor Model (DFM):

\[ X_t = \Lambda f_t + v_t \]
\[ f_t = \Phi(L)f_{t-1} + H\varepsilon_t \]
\[ v_{i,t} = \Delta_i(L)v_{i,t-1} + \Xi_i\xi_{i,t} \]

• \( X_t \): 207 quarterly macro time series, spanning various categories.

• \( f_t \): six latent driving factors, evolve as VAR(2), driven by six aggregate shocks \( \varepsilon_t \).

• \( v_{i,t} \): idiosyncratic noise, evolves as AR(2), independent across \( i \).

• Parameters fixed at estimates from U.S. data in Stock & Watson (2016). (\( H \): next slide.)

• Gaussian shocks + noise.
Lower-dimensional DGPs and estimands

- Draw 6,000 subsets of 5 variables $\tilde{w}_t \subset X_t$. DFM implies that $\tilde{w}_t$ follows $\text{VAR}(\infty)$.
- $\tilde{w}_t$ contains at least one output and one price series, and – depending on type of DGP...
  1. Monetary shock: $i_t =$ federal funds rate.
  2. Fiscal shock: $i_t =$ federal government spending.
- Select response variable $y_t \in \tilde{w}_t$ at random (not $i_t$).
- For today, assume shock $\varepsilon_{1,t}$ is observed. Estimand: $\theta_h = \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}}$, $h = 0, 1, 2, \ldots, 19$.
  - Other ID schemes in paper: recursive, proxy/IV.
- $H = \frac{\partial f_t}{\partial \varepsilon_t}$ chosen to maximize impact response of $i_t$ wrt. $\varepsilon_{1,t}$. 
DGPs and impulse response estimands are heterogeneous
1. Analytical example
2. Data generating processes
3. Estimators
4. Results
5. Conclusion
Impulse response estimators

- Local projection methods:
  1. **Least squares.** Jordà (2005)
  2. **Penalized**: shrinks towards quadratic polynomial in $h$. Barnichon & Brownlees (2019)

- VAR methods:
  3. **Least squares.**
  4. **Bias-corrected**: corrects small-sample bias due to persistence. Pope (1990)
  5. **Bayesian**: Minnesota-type prior, shrinks towards white noise. Canova (2007)
  6. **Model averaging**: Data-dependent weighted average of estimates from 40 models, AR(1) to AR(20) and VAR(1) to VAR(20). Hansen (2016); Miranda-Agrippino & Ricco (2019)
1. Analytical example
2. Data generating processes
3. Estimators
4. Results
5. Conclusion
Specification and simulation settings

- Include $p = 4$ lags in LP and VAR, except VAR model averaging ($p = 8$ in paper). AIC almost always selects fewer than 4 lags.

- Show results for 6,000 monetary and fiscal shock DGPs jointly (separate results in paper).

- Loss function:

$$L_\omega(\theta_h, \hat{\theta}_h) = \omega \times (\mathbb{E}[\hat{\theta}_h - \theta_h])^2 + (1 - \omega) \times \text{Var}(\hat{\theta}_h).$$

Divide estimator bias/std by $\sqrt{\frac{1}{20} \sum_{h=0}^{19} \theta_h^2}$ to remove units.

- $T = 200$. 5,000 Monte Carlo repetitions per DGP.
  - Simulations take two weeks on the cluster in Matlab with 25 parallel cores.
Lesson 1: There is a clear bias-variance trade-off between LP and VAR

Observed shock identification, medians across 6,000 DGPs
Lesson 2: Shrinkage dramatically lowers variance, at some cost of bias

**Observed shock identification, medians across 6,000 DGPs**
Lesson 2: Shrinkage dramatically lowers variance, at some cost of bias.

Observed shock identification, medians across 6,000 DGPs.
Lesson 3: No method dominates, but shrinkage is generally welcome

Observed shock: Average loss minimizing estimator
Can we select the estimator based on the data?

- In-sample, data-driven estimator choice $\implies$ best of both worlds?
- Disappointing performance of VAR model averaging estimator suggests caution.
- In our DGPs, conventional model selection/evaluation criteria are unable to detect even substantial misspecification of VAR(4) model.
  - AIC: 90th percentile of $\hat{p}_{AIC}$ does not exceed 2 in any DGP.
  - LM test of residual serial correlation: rejection probability below 25% in 99.9% of DGPs (signif. level = 10%).
Outline

1. Analytical example
2. Data generating processes
3. Estimators
4. Results
5. Conclusion
Conclusion

- Large-scale simulation study of LP, VAR, and related impulse response estimators.
- Thousands of DGPs drawn from encompassing empirical DFM.
- Lessons:
  1. Clear bias-variance trade-off between least-squares LP and VAR. Loss fct weight on bias must be high to prefer LP over VAR. Caveat: moderately persistent DGPs.
  2. Shrinkage dramatically lowers variance, at some cost of bias.
  3. No method dominates at all horizons, but shrinkage is generally welcome. Penalized LP good at short horizons, BVAR good at intermediate+long.
- In paper: recursive and proxy/IV identification.
Conclusion

- Large-scale simulation study of LP, VAR, and related impulse response estimators.
- Thousands of DGPs drawn from encompassing empirical DFM.
- Lessons:
  1. Clear bias-variance trade-off between least-squares LP and VAR. Loss fct weight on bias must be high to prefer LP over VAR. Caveat: moderately persistent DGPs.
  2. Shrinkage dramatically lowers variance, at some cost of bias.
  3. No method dominates at all horizons, but shrinkage is generally welcome. Penalized LP good at short horizons, BVAR good at intermediate+long.
- In paper: recursive and proxy/IV identification.

Thank you!
• Direct vs. iterated forecasts:
  Marcellino, Stock & Watson (2006)

• LP vs. VAR simulation studies:
  Jordà (2005); Meier (2005); Kilian & Kim (2011); Brugnolini (2018); Choi & Chudik (2019); Austin (2020); Bruns & Lütkepohl (2021)

• Analytical comparisons:
  Schorfheide (2005); Kilian & Lütkepohl (2017); Plagborg-Møller & Wolf (2021)

• Shrinkage estimation:
  Hansen (2016); Barnichon & Brownlees (2019); Miranda-Agrippino & Ricco (2019)

• Inference about IRFs:
  Inoue & Kilian (2020); Montiel Olea & Plagborg-Møller (2020)
Proposition 1

Fix $h \geq 0$, $\rho \in (-1, 1)$, $\sigma_2 > 0$, and $\alpha \in \mathbb{R}$. Assume $E(\varepsilon_{j,t}^4) < \infty$ for $j = 1, 2$. Define $\sigma_{0,y}^2 \equiv \frac{1+\sigma_2^2}{1-\rho^2}$. Then, as $T \to \infty$,

$$
\sqrt{T}(\hat{\beta}_h - \theta_h) \overset{d}{\to} N(\text{aBias}_{\text{LP}}, \text{aVar}_{\text{LP}}), \quad \sqrt{T}(\hat{\delta}_h - \theta_h) \overset{d}{\to} N(\text{aBias}_{\text{VAR}}, \text{aVar}_{\text{VAR}}),
$$

where for all $h \geq 0$,

$$
a\text{Bias}_{\text{LP}} \equiv 0, \quad a\text{Var}_{\text{LP}} \equiv \sigma_{0,y}^2(1 - \rho^{2(h+1)}) - \rho^{2h},
$$

and for $h \geq 1$,

$$
a\text{Bias}_{\text{VAR}} \equiv \rho^{h-1}(h-1)\frac{\alpha\sigma_2^2}{\sigma_{0,y}^2 - 1}, \quad a\text{Var}_{\text{VAR}} \equiv \rho^{2(h-1)}(1 - \rho^2)\sigma_{0,y}^2 \left(1 + \frac{(h-1)^2}{\sigma_{0,y}^2 - 1}\right) + \rho^{2h}\sigma_2^2.
$$
Interpretation of degree $\alpha$ of misspecification

**Proposition 2**

Impose same assumptions as in Proposition 1.

Let $\hat{\tau}$ denote the t-statistic for testing the significance of the second lag in a univariate AR(2) regression for $\{y_t\}$.

Then, as $T \to \infty$,

$$ \hat{\tau} \overset{d}{\to} N \left( -\rho \frac{\sigma_2^2}{1 + \sigma_2^2} \alpha, 1 \right). $$
DGPs are heterogeneous along various dimensions

<table>
<thead>
<tr>
<th>Percentile</th>
<th>min</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data and shocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trace(long-run var)/trace(var)</td>
<td>0.42</td>
<td>0.93</td>
<td>0.98</td>
<td>1.14</td>
<td>2.29</td>
<td>4.78</td>
<td>18.09</td>
</tr>
<tr>
<td>Largest VAR eigenvalue</td>
<td>0.82</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td>Fraction of VAR coef’s $\ell \geq 5$</td>
<td>0.02</td>
<td>0.10</td>
<td>0.15</td>
<td>0.23</td>
<td>0.34</td>
<td>0.44</td>
<td>0.84</td>
</tr>
<tr>
<td>Degree of shock invertibility</td>
<td>0.14</td>
<td>0.16</td>
<td>0.19</td>
<td>0.28</td>
<td>0.41</td>
<td>0.47</td>
<td>0.65</td>
</tr>
<tr>
<td>IV first stage F-statistic</td>
<td>9.49</td>
<td>9.61</td>
<td>9.70</td>
<td>17.29</td>
<td>27.58</td>
<td>28.20</td>
<td>29.10</td>
</tr>
<tr>
<td><strong>Impulse responses up to $h = 19$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of interior local extrema</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Horizon of max abs. value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Average/(max abs. value)</td>
<td>-0.44</td>
<td>-0.16</td>
<td>-0.09</td>
<td>-0.02</td>
<td>0.06</td>
<td>0.11</td>
<td>0.46</td>
</tr>
<tr>
<td>$R^2$ in regression on quadratic</td>
<td>0.01</td>
<td>0.11</td>
<td>0.23</td>
<td>0.49</td>
<td>0.71</td>
<td>0.84</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Combining 6,000 monetary and fiscal DGPs. Observed shock or IV identification.
Example IRF estimates: Least-squares LP
Example IRF estimates: Penalized LP
Example IRF estimates: Least-squares VAR

![Graph showing impulse response functions for a least-squares VAR model. The graph plots the response of one variable to a shock in another variable over various horizons.]
Example IRF estimates: Bias-corrected VAR
Example IRF estimates: Bayesian VAR
Example IRF estimates: VAR model averaging
Lesson 1: There is a clear bias-variance trade-off between LP and VAR.
Lesson 2: Shrinkage dramatically lowers variance, at some cost of bias.

LP preferred over Pen LP

VAR preferred over BVAR

Observed shock identification
Lesson 4: SVAR-IV is heavily biased, but has relatively low dispersion

**Median bias**

**Interquartile range**

IV identification, medians across 6,000 DGPs