Misspecification-Robust Shrinkage and Selection for VAR Forecasts and IRFs

Universitat Pompeu Fabra Barcelona

Oriol González-Casasús[1] Frank Schorfheide[2]

[1]UPF & BSE [2]UPenn, CEPR, NBER & PIER

BSE Barcelona School of Economics

Introduction

VARs popular for forecasting and IRF estimation, but...

- Large systems: observations/parameter ratio is small.
- Potential dynamic misspecification.

Popular empirical strategies:

- Bayesian estimation with shrinkage priors (e.g., Minnesota prior).
- Direct (LFE/LP) vs. Iterated (MLE/VAR) approaches.

This paper:

- Criteria to simultaneously select shrinkage (hyperparameters), lag order and type of estimator.
- Tailored to forecasting and IRF applications.
- Robust to misspecification.

Framework

Model and DGP

- **Data generating process**, local to VAR(p_*), $0 \le p_* < q$:

$$y_t = F_1 y_{t-1} + \cdots + F_{p_*} y_{t-p_*} + \epsilon_t + \frac{\alpha}{\sqrt{T}} \sum_{j=1}^{\infty} A_j \epsilon_{t-j}.$$

- **Econometrician's model**, VAR(p), $0 \le p \le q$:

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t.$$

Estimands of Interest ($p = 1$)

1. **Likelihood-Based (MLE/VAR)**: Posterior mean based on

$$y_t = \Phi y_{t-1} + u_t, \quad u_t \sim N(0, \Sigma_{uu}), \quad \Phi|\Sigma_{uu} \sim N\big(\underline{\Phi}_T,\ (\tilde{\lambda}_T \underline{P}_\Phi)^{-1} \otimes \Sigma_{uu}\big).$$

2. **Loss-Function-Based (LFE/LP)**: Posterior mean based on

$$y_t = \Psi y_{t-h} + v_t, \quad v_t \sim N(0, \Sigma_{vv}), \quad \Psi|\Sigma_{vv} \sim N\big(\underline{\Psi}_T,\ (\tilde{\lambda}_T \underline{P}_\Psi)^{-1} \otimes \Sigma_{vv}\big).$$

Drifting Priors

- Mean in a $T^{-1/2}$ radius of F with precision $\tilde{\lambda}_T = T\lambda$.
- Shrinkage weights are non-trivial even asymptotically.

Main Goal

Posterior mean $\bar{\Psi}_T(\iota, \lambda, p)$ used to construct forecasts/IRFs depends on:

- ι: type of estimand (MLE/VAR or LFE/LP).
- λ: degree of shrinkage ($\mathbb{R}_+$).
- p: autoregressive lag order $(0, 1, \ldots, q)$.
- ▷ Horizon h implicitly fixed.

⇒ How to select (ι, λ, p) to minimize risk?

Limit Distribution of Shrinkage Estimators

Theorem. *Under regularity conditions, for* $\iota \in \{lfe, mle\}$, $\lambda \ge 0$, *and* $p \le q$:

$$\bar{\Psi}_T(\iota, \lambda, p) = F^h + T^{-1/2}[\delta(\iota, \lambda, p) + \alpha\mu(\iota, \lambda, p) + \zeta_T(\iota, \lambda, p)] + o_p(T^{-1/2}).$$

Moreover, $\zeta_T(\iota, \lambda, p) \xrightarrow{d} N(0, V(\iota, \lambda, p))$.

- Two sources of bias: shrinkage (δ) and misspecification (μ).
- Noise term: ζ_T.

Model Determination: Forecasting

The h-step-ahead forecast is

$$\hat{y}_{T+h}(\iota, \lambda, p) = \bar{\Psi}_T(\iota, \lambda, p) y_T,$$

with asymptotic risk $\bar{\mathcal{R}}(\hat{y}_{T+h}(\iota, \lambda, p))$.

Theorem. *Define*

$$PC_T(\iota, \lambda, p) = T[W \cdot MSE(\iota, \lambda, p)] + \text{Pen}(\iota, \lambda, p).$$

Under regularity conditions,

$$\mathbb{E}[PC_T(\iota, \lambda, p) - PC_T(\iota', \lambda', p')] \to \bar{\mathcal{R}}(\hat{y}_{T+h}(\iota, \lambda, p)) - \bar{\mathcal{R}}(\hat{y}_{T+h}(\iota', \lambda', p')).$$

⇒ Forecasting model determination based on PC_T.

Model Determination: IRF

The IRF of a shock ϵ_{t-h} on y_t is

$$\frac{\partial y_t}{\partial \epsilon'_{t-h}} = F^h + \frac{\alpha}{\sqrt{T}} \mu(irf, p), \qquad \mu(irf, p) = \sum_{j=0}^{h-1} F^j A_{h-j},$$

with asymptotic risk $\bar{\mathcal{R}}_{IRF}(\iota, \lambda, p)$ and shock identification Ξ.

Theorem. *Define*

$$IRFC_T(\iota, \lambda, p) = T\|\bar{\Psi}_T(\iota, \lambda, p) - \bar{\Psi}(lfe, 0, q)\|^2_{W \otimes \Xi\Xi'} + \text{Pen}(\iota, \lambda, p).$$

Under regularity conditions,

$$\mathbb{E}[IRFC_T(\iota, \lambda, p) - IRFC_T(\iota', \lambda', p')] \to \bar{\mathcal{R}}_{IRF}(\iota, \lambda, p) - \bar{\mathcal{R}}_{IRF}(\iota', \lambda', p').$$

⇒ IRF model determination based on $IRFC_T$.

- Key to unbiasedness is lag-augmentation ($q > p^*$).

Monte Carlo Experiment: Forecasting

- Calibrate VAR(1) to Carriero, Clark and Marcellino (2015), with VMA(10) drift.
- Local LFE prior mean $\underline{\psi} = \varphi\mu(pov)$, and MLE prior aligned such that

$$\underline{\psi} = \sum_{j=0}^{h-1} F^j \underline{\phi} F^{h-1-j}.$$

- $(\alpha, \varphi) = (0, 1)$ and $(\alpha, \varphi) = (2, 0.5)$. Here, results for $h = 4$, $T = 500$.

Risk at $\hat{\lambda}$

	$\alpha = 0$				$\alpha = 2$			
p	LFE	MLE	$\hat{\iota}$	% LFE	LFE	MLE	$\hat{\iota}$	% LFE
1	-194	**-197**	-195	39	**39**	51	40	90
2	-190	**-196**	-193	30	**25**	43	26	96
4	-185	**-191**	-189	27	**-40**	10	-40	100
6	-182	**-187**	-185	27	**-58**	-33	-58	97
$\hat{p}$	-188	**-191**	-189	25	**-57**	-32	-57	99

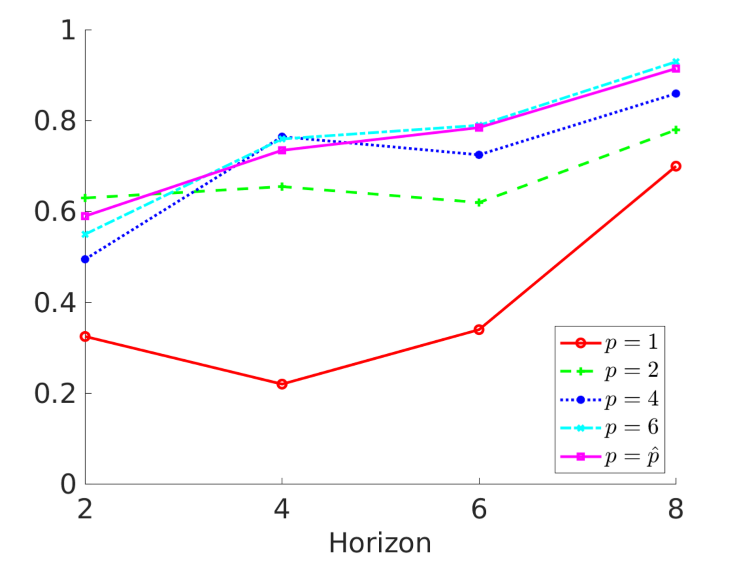
PC vs. MDD

	$\alpha = 0$		$\alpha = 2$	
p	LFE	MLE	LFE	MLE
1	**-7**	0	**-13**	**-18**
2	0	4	34	**-3**
4	5	0	**-368**	**-86**
6	5	1	**-181**	**-143**
$\hat{p}$	**-3**	4	**-226**	**-152**

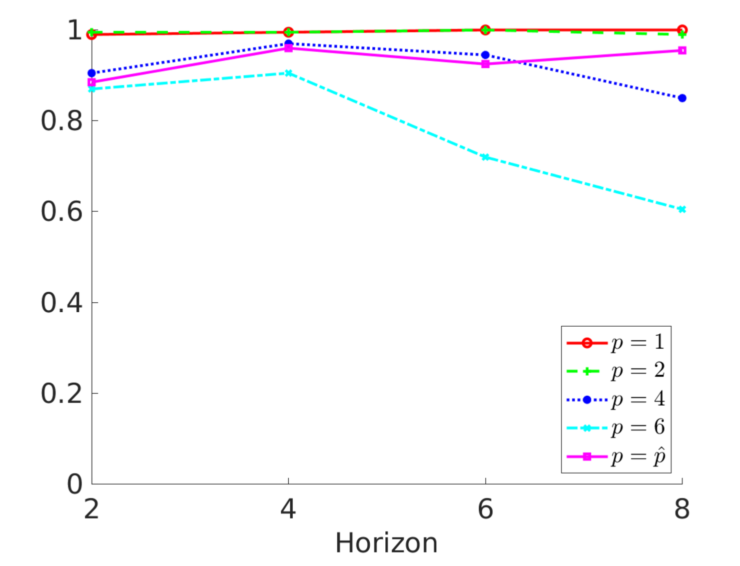
Empirical Application: IRF Estimation

- How frequently LPs or VARs? How much shrinkage when estimating IRFs?
- FRED-QD database, transformed to stationary via Hamilton (2018).
- Create a large number of datasets by randomly selecting 200 different six-tuples of series. (Marcellino, Stock and Watson, 2006)
- $h = 6$, 1984:Q1-2006:Q4, $\Xi = \hat{\Sigma}_{chol}$.

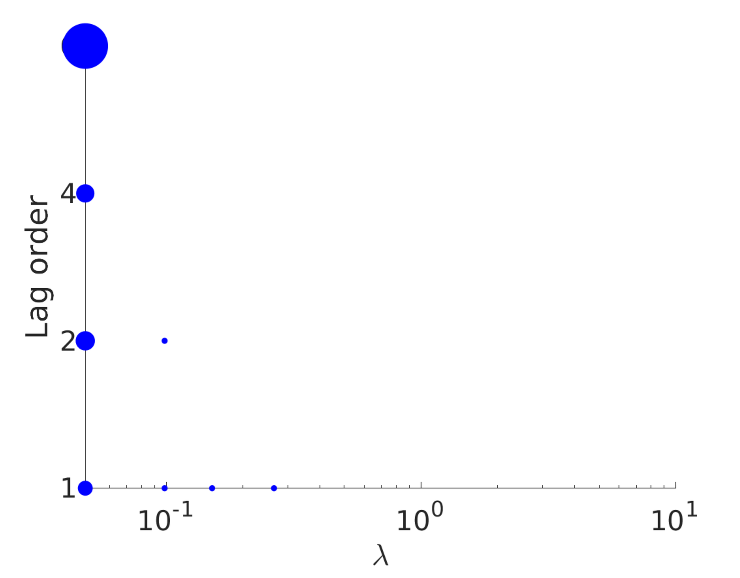
%LP (IRFC)



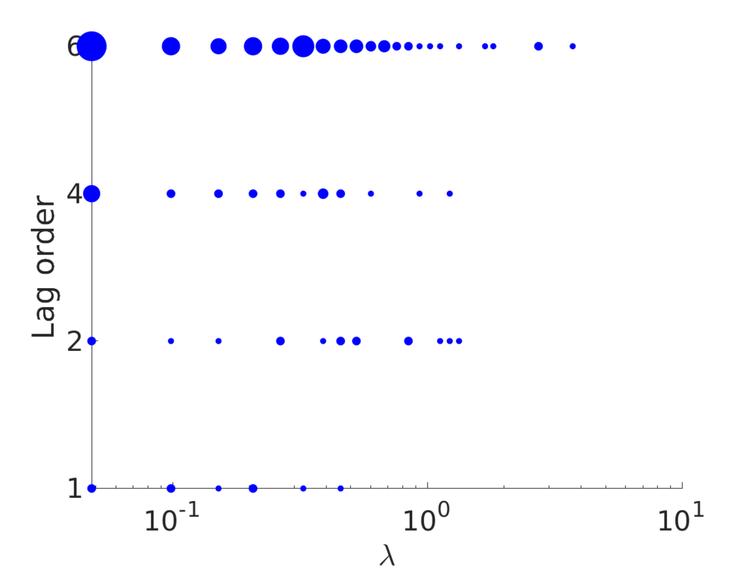
%LP (PC)



Local Projection



VAR-Based



References

Carriero, A., Clark, T. E. and Marcellino, M. (2015), 'Bayesian vars: Specification choices and forecast accuracy', *Journal of Applied Econometrics* 30(1), 46-73.

Hamilton, J. D. (2018), 'Why you should never use the Hodrick-Prescott filter', *Review of Economics and Statistics* 100(5), 831-843.

Marcellino, M., Stock, J. H. and Watson, M. W. (2006), 'A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series', *Journal of Econometrics* 135, 499-526.

13th ECB Conference on Forecasting Techniques