Forecasts in a Slightly Misspecified Finite Order VAR

Ulrich K. Müller
Princeton University

James H. Stock
Harvard University

June 2011

Introduction

- Benchmark approach for statistical forecasting: Forecast from AR(p)

$$y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + e_t$$

  with AR parameters estimated by OLS and $p = \hat{p}_{BIC}$ determined by Bayesian Information Criterion (BIC).

- Perfectly reasonable (=admissible up to $o_p(T^{-1/2})$ error) if true DGP is Gaussian AR(p_0), as $\hat{p}_{BIC} \xrightarrow{p} p_0$, and OLS is equal to MLE.

- But not otherwise.

- AIC selects larger p, $\hat{p}_{AIC} > \hat{p}_{BIC}$:

  - $\hat{p}_{AIC} > p_0$ with positive asymptotic probability under AR(p_0)

  - AIC has some asymptotic optimality in class of AR(p) forecasts when DGP is AR(∞) (Shibata (1980), Schorfheide (2005), Ing and Wei (2005)), but might well be inadmissible overall

Are Macroeconomic Time Series AR(p_0)'s?

Consider difference $\hat{p}_{AIC} - \hat{p}_{BIC}$ in the 132 Stock and Watson (2005) monthly macro U.S. postwar time series, and compare to asymptotics under AR(p_0)

$\hat{p}_{AIC} - \hat{p}_{BIC}$	Empirical	Asymptotic
0	0.205	0.713
1	0.098	0.113
2	0.083	0.055
3	0.098	0.033
4	0.068	0.023
5	0.030	0.016
6	0.038	0.011
7	0.045	0.008
8	0.045	0.006
9	0.030	0.005
10	0.015	0.003
11	0.053	0.003
$>$12	0.189	0.010

Local-To-Flat Spectral Density of e_t

Model Gaussian errors e_t in baseline AR(p_0) model $\beta(L)y_t = e_t$ as slightly predictable: Spectral density is

$$f_e(\omega) = \frac{1}{2\pi} e^{G(\omega)/\sqrt{T}}$$

for some non-constant function G satisfying $\int_{-\pi}^{\pi} G(\omega) d\omega = 0$.

- $\int_{-\pi}^{\pi} G(\omega) d\omega = 0$ implies $V[e_t | e_{t-1}, e_{t-2}, \ldots] = 1$.

- But unconditional variance is

$$\begin{aligned} V[e_t] &= \int_{-\pi}^{\pi} f_e(\omega) d\omega \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1 + G(\omega)/\sqrt{T} + \frac{1}{2} G(\omega)^2 / T\right) d\omega \\ &= 1 + T^{-1} \frac{1}{4\pi} \int_{-\pi}^{\pi} G(\omega)^2 d\omega. \end{aligned}$$

$\Rightarrow$ Ignoring predictability in e_t leads to $O_p(T^{-1/2})$ error in forecast, same order of magnitude as parameter uncertainty about β.

Inference in Slightly Misspecified AR(p)

- AR(p_0) model with local-to-flat $f_e(\omega)$ is *contiguous* to model "pure" AR(p_0) model with $f_e(\omega) = 1/2\pi$: Impossible to consistently estimate G

- Bayesian approach: Treat G as realization of a demeaned Gaussian process
$$G(\omega) = J(\omega) - \frac{1}{\pi}\int_0^\pi J(r)dr.$$

- Example I: Let
$$J \sim cW$$
where W is standard Wiener process and scalar c determines the degree of non-flatness.

- Example II: Let
$$J(\omega) = cW_{\bar{\omega}}(\omega) = \begin{cases} \frac{c}{\sqrt{\bar{\omega}}}W(\omega) & \text{for } 0 \leq \omega < \bar{\omega} \\ \frac{c}{\sqrt{\bar{\omega}}}W(\bar{\omega}) & \text{otherwise} \end{cases}$$

AIC and BIC in Model with $J \sim cW$

$\hat{p}_{AIC} - \hat{p}_{BIC}$	Empirical	Asymptotic $c = 0$	Asymptotic $p_0 = 0$, $c = 20$	Asymptotic $p_0 = 3$, $c = 40$
0	0.205	0.713	0.185	0.165
1	0.098	0.113	0.174	0.115
2	0.083	0.055	0.153	0.107
3	0.098	0.033	0.123	0.096
4	0.068	0.023	0.092	0.088
5	0.030	0.016	0.068	0.074
6	0.038	0.011	0.051	0.065
7	0.045	0.008	0.037	0.053
8	0.045	0.006	0.027	0.044
9	0.030	0.005	0.021	0.037
10	0.015	0.003	0.015	0.029
11	0.053	0.003	0.012	0.025
$>$12	0.189	0.010	0.041	0.102

Forecasts with Gaussian Process Prior on G

- Under squared loss, best forecast of e_{T+1} is posterior mean.

- Literature on Bayesian estimation of time series models with priors in spectral domain: Carter and Kohn (1997), etc.

  - Computationally intensive

  - Posterior samplers are based on Whittle approximation to likelihood $\Rightarrow$ induces $O_p(T^{-1/2})$ error in posterior mean

- Main contribution of the paper: Computationally convenient, closed-form and to order $o_p(T^{-1/2})$ accurate approximation to posterior mean

Heuristics of Approximation

- With $f_e(\omega) = \frac{1}{2\pi} e^{G(\omega)/\sqrt{T}}$, $\gamma_j(G) = \sqrt{T} E[e_t e_{t-j}]$ is approx. linear in G

$$\gamma_j(G) = 2\sqrt{T} \int_0^\pi \cos(\omega k) f_e(\omega) d\omega \approx \frac{1}{\pi} \int_0^\pi \cos(\omega k) G(\omega) d\omega$$

  so Gaussian prior on G implies approximately Gaussian prior on $\{\gamma_j(G)\}_{j=1}^T$.

- Local flatness allows for accurate quadratic approximation to log-likelihood in $\{\gamma_j\}_{j=1}^T$, with mean $\hat{\gamma}_j = T^{-1/2} \sum_{t=1}^T e_t e_{t-j}$, so that likelihood information about $\gamma_j(G)$ is also approximately Gaussian.

- Let $e = (e_T, e_{T-1}, \cdots, e_1)'$. Then

$$\begin{aligned} \sqrt{T} E[e_{T+1} | G, e] &= \sqrt{T} E[e_{T+1} e' | G] V[e|G]^{-1} e \\ &\approx (\gamma_1(G), \gamma_2(G), \cdots, \gamma_{T+1}(G)) e \end{aligned}$$

$\Rightarrow$ optimal forecast of e_{T+1} given $\{e_t\}_{t=1}^T$ is approximately equal to mean of posterior in a Gaussian–Gaussian prior–likelihood problem.

Details on Quadratic Log-Likelihood in $\gamma(G)$

- With $e \sim \mathcal{N}(0, V(G))$, $V(G) = I + T^{-1/2}\Gamma(G)$

$$l(G) = -\tfrac{1}{2}\ln\det V(G) - \tfrac{1}{2}e'V(G)^{-1}e$$

- But $V(G)^{-1} \approx I - T^{-1/2}\Gamma(G) + T^{-1}\Gamma(G)^2$, so that

$$\begin{aligned} -\tfrac{1}{2}e'V(G)^{-1}e &\approx -\tfrac{1}{2}e'e + \tfrac{1}{2}T^{-1/2}e'\Gamma(G)e - \tfrac{1}{2}T^{-1}e'\Gamma(G)^2e \\ &\approx C + \hat{\gamma}'\gamma(G) - \tfrac{1}{2}T^{-1}\operatorname{tr}\Gamma(G)^2 \end{aligned}$$

and since $\ln(1+x) \approx x - \frac{1}{2}x^2$

$$\begin{aligned} -\tfrac{1}{2}\ln\det V(G) &= -\tfrac{1}{2}\sum_{i=1}^{T}\ln(1+\lambda_{T,i}(G)) \\ &\approx -\tfrac{1}{2}T^{-1/2}\operatorname{tr}\Gamma(G) + \tfrac{1}{4}T^{-1}\operatorname{tr}\Gamma(G)^2 \end{aligned}$$

so that with $T^{-1/2}\operatorname{tr}\Gamma(G) \approx \frac{1}{\pi}\int_0^\pi G(\omega)d\omega = 0$,

$$l(G) \approx \hat{\gamma}'\gamma(G) - \tfrac{1}{4}T^{-1}\operatorname{tr}\Gamma(G)^2 \approx \hat{\gamma}'\gamma(G) - \tfrac{1}{2}\gamma(G)'\gamma(G).$$

Approximation with AR(p) Baseline

- The OLS estimates $\{\hat{\beta}_j\}_{j=1}^p$ are large sample equivalent to posterior mean of $\{\beta_j\}_{j=1}^p$ under any non-dogmatic prior.

- But spectral density of

$$\hat{e}_t = \hat{\beta}(L)y_t = \hat{\beta}(L)\beta(L)^{-1}e_t$$

  is not the same as that of e_t up to order $O_p(T^{-1/2})$.

- Adjustment for the presence of $\hat{\beta}(L)\beta(L)^{-1}$ to obtain approximately optimal forecast of $\hat{e}_{T+1}$ given $\{\hat{e}_t\}_{t=1}^T$.

- Approximate forecast of $y_{T+\ell}$ is simply obtained by iterating $\hat{\beta}(L)y_t = \hat{e}_t$ forward with future $\hat{e}_t$ set to their approximate posterior means.

Formal Result

$k \times 1$ VAR generalization

$$
\begin{aligned}
y_t &= \alpha + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + P e_t \\
f_e(\omega) &= \frac{1}{2\pi} \exp[T^{-1/2} G(\omega)]
\end{aligned}
$$

Theorem:

(a) For a wide class of Gaussian process priors on G, and a non-dogmatic prior on (α, β), the approximation error of the approximate forecast of $y_{T+\ell}$ is $o_p(T^{-1/2})$ for any fixed forecast horizon ℓ.

(b) Computationally straightforward approximation to Bayesian model averaging (BMA) weights for prior that is a discrete mixture of different G processes are consistent for the exact BMA weights (so that posterior mean approximation error remains $o_p(T^{-1/2})$ under mixture prior).

Small Sample Results

- Three questions for forecasting rules constructed from $J \sim cW$

  1. How good is the approximation relative to exact posterior mean forecast?
  2. How well does the approximate BMA do?
  3. How good are the approximate forecasts when true model is MA(1)?

- Report numbers in terms of

$$\frac{T(\text{MSFE} - \text{MSFE}_{\text{AR}(p)})}{\text{MSFE}_{\text{AR}(p)}}$$

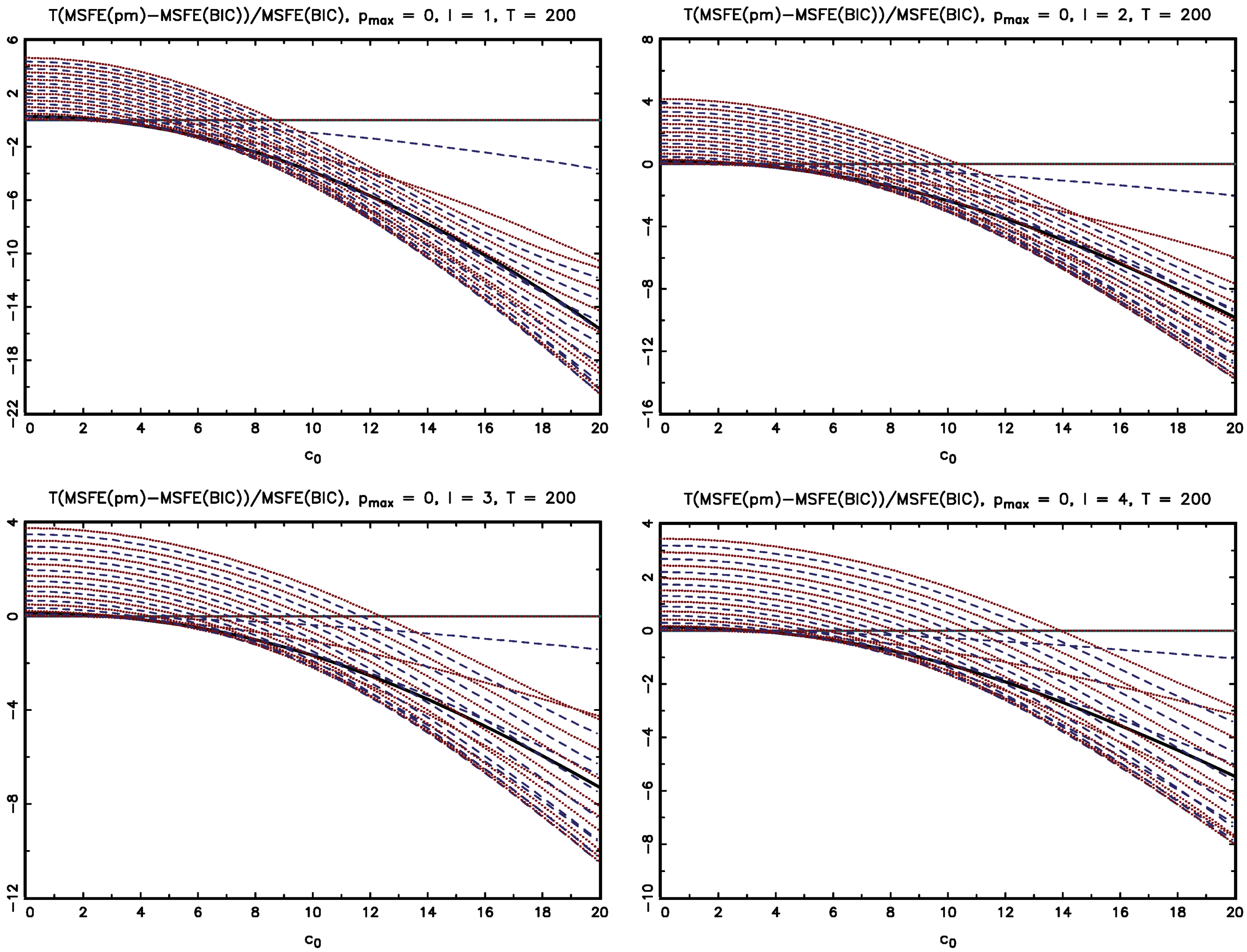
$\Rightarrow$ Forecasting an AR(p_0) with an OLS estimate of AR(p) with $p > p_0$ instead of OLS estimate of AR(p_0) induces deterioration of $p - p_0$ in this rescaled MSFE measure

Small Sample Approximation Quality

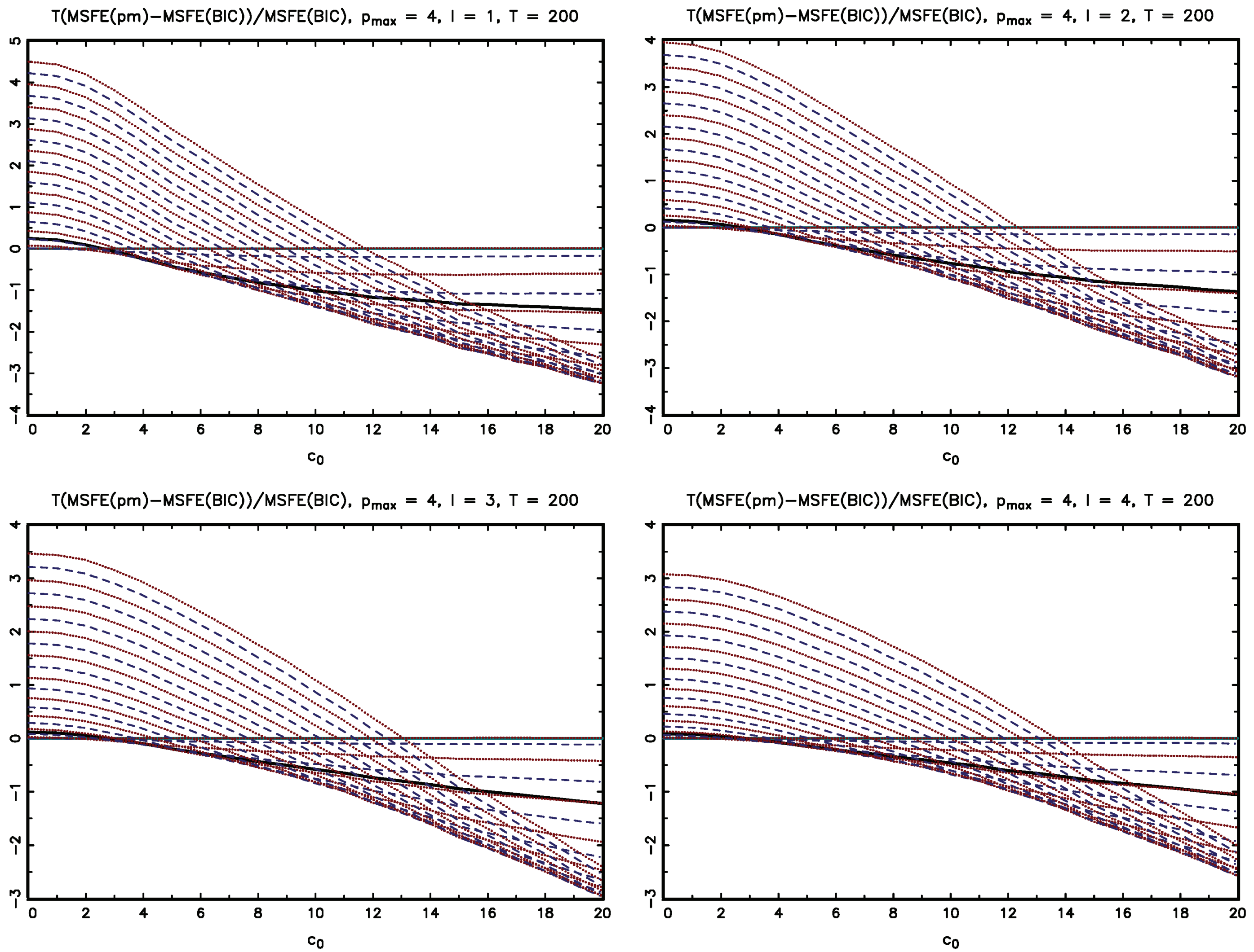
DGP and prior have $J \sim cW$

$\frac{T(\text{MSFE}-\text{MSFE}_{\text{AR}(p)})}{\text{MSFE}_{\text{AR}(p)}}$	$c = 10$			$c = 30$		
	$p = 0$	$p = 1$	$p = 3$	$p = 0$	$p = 1$	$p = 3$
			$T = 100$			
Exact	-4.6	-1.0	-0.4	-49.1	-22.6	-6.6
Approximate	-4.1	-1.0	-0.2	31.9	-8.6	-4.9
Difference	-0.6	0.0	-0.2	-81.1	-13.9	-1.7
			$T = 200$			
Exact	-5.4	-0.9	-0.4	-55.7	-21	-6.8
Approximate	-5.2	-0.9	-0.3	-14.6	-14.2	-6.0
Difference	-0.2	0.0	0.0	-41.1	-6.8	-0.8
			$T = 400$			
Exact	-6.2	-1.0	0.0	-64.8	-20.9	-6.4
Approximate	-6.1	-1.0	0.0	-39.9	-17.5	-6.3
Difference	-0.1	0.0	0.0	-24.9	-3.3	-0.2

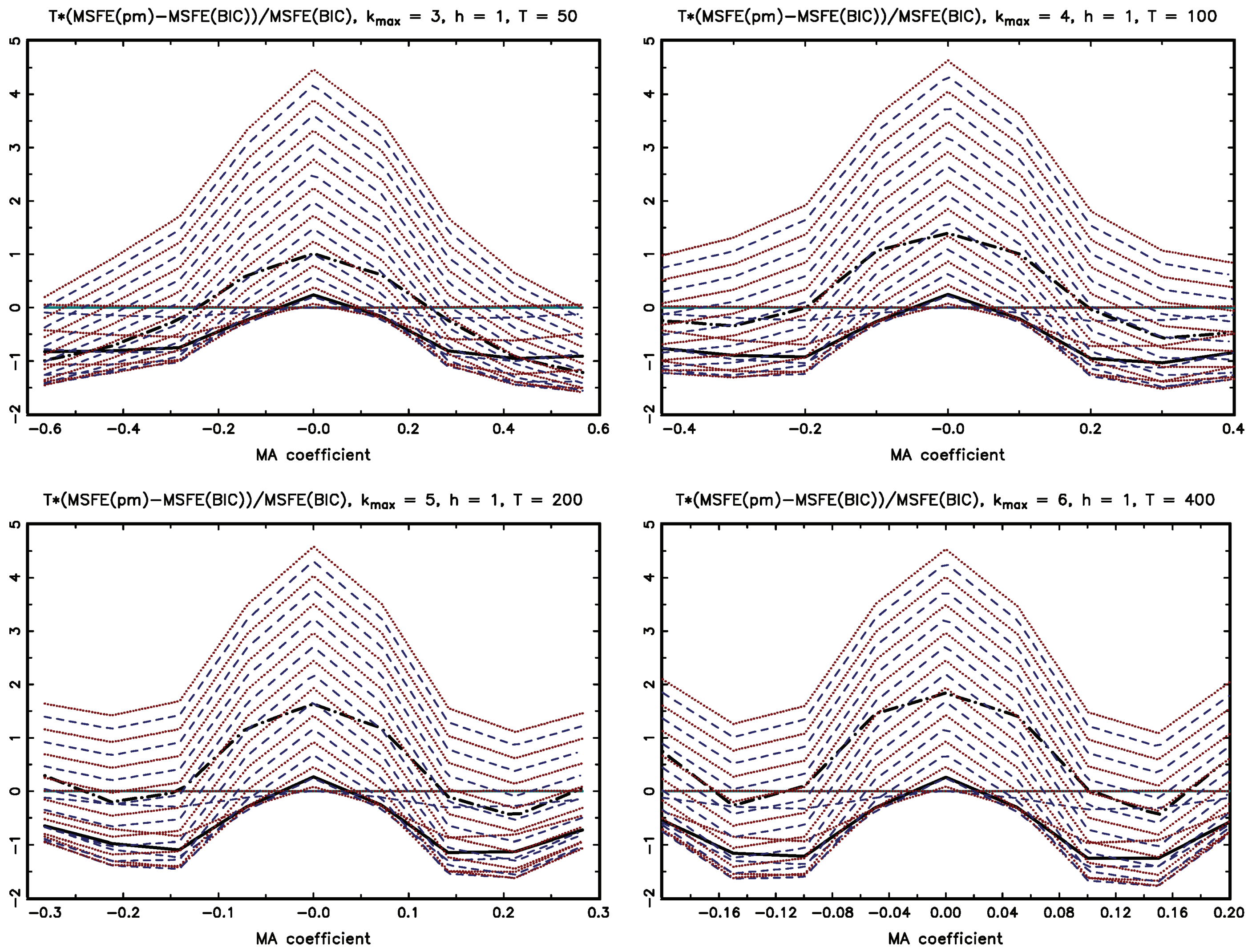
Envelope and BMA, AR(0) Imposed



Envelope and BMA, AR(p_{BIC}), $p_{\max} = 4$



Performance in MA(1) Model



Empirical Performance

MSFE relative to BIC in pseudo out-of-sample recursive AR estimates in 132 monthly post war series from Stock and Watson (2005)

		$J \sim cW$			$J \sim cW + 20W_{2\pi/96}$		
Horizon ℓ	AIC	$c = 20$	$c = 30$	BMA	$c = 20$	$c = 30$	BMA
				Mean			
1	1.025	0.988	0.988	0.989	0.987	0.988	0.988
6	0.982	0.968	0.963	0.97	0.963	0.958	0.966
				Median			
1	1.002	0.988	0.988	0.989	0.987	0.987	0.988
6	0.997	0.98	0.976	0.982	0.974	0.969	0.976
				10% Percentile			
1	0.956	0.964	0.953	0.964	0.960	0.951	0.962
6	0.858	0.909	0.888	0.925	0.900	0.880	0.913
				90% Percentile			
1	1.135	1.014	1.023	1.011	1.012	1.020	1.012
6	1.091	1.008	1.015	1.006	1.015	1.023	1.016

Conclusion

- Convenient method to exploit residual predictability in disturbances of parsimonious VARs
- Could use local-to-flat spectral framework for purposes other than forecasting
  - Approximate BMA weights can be used as asymptotic weighted average power maximizing specification tests for VAR(p)
  - Study effect of local misspecification in spectral domain for inference about VAR parameters, impulse responses, etc.