Neglecting Parameter Changes in Autoregressive Models

Eric Hillebrand
Louisiana State University

Midwest Econometrics Group Meeting,
Carbondale, IL, Oct. 15, 2005

THE PHENOMENON

- Autoregressive models estimated on data containing mean shifts

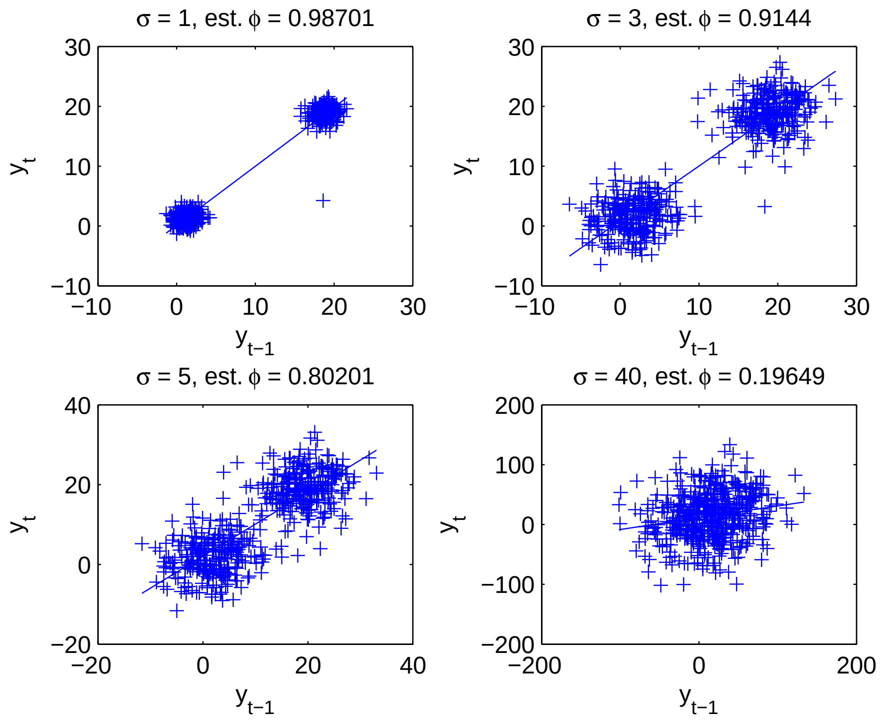
$$\Longrightarrow \sum_{j=1}^{p} \widehat{\phi}_j \approx 1$$

.

- Cause of mean shifts: Changes in constant or autoregressive parameter.
- Effects of mean shifts: Persistence and impulse response estimation.
- GARCH: $\widehat{\alpha} + \widehat{\beta} \approx 1$, no matter what type of data?

GEOMETRY OF THE PROBLEM

There is a trade-off between the size of the breaks and the variance of the stochastic driver of the process. (DGP AR(1) with $\phi = 0.20$, break occurs at $T/2 = 200$.)



RELATED WORK

- Perron 1989, Hendry and Neale 1991 (power of Dickey-Fuller tests)
- Perron 1990, Bai 1994, 1997, Bai and Perron 1998, Kokoszka and Leipus 1999, Berkes et al. 2004 (change-point detection)
- Chen and Tiao 1990, Mikosch and Starica 2004, Ng and Vogelsang 2002 (neglected parameter changes)

MODEL

The data generating process is

$$\mathbf{y}_t = \begin{cases} \mathbf{c}_1 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, & t \in \underline{T_1} \\ \dots \\ \mathbf{c}_K + \mathbf{\Phi}_K \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, & t \in \underline{T_K}, \end{cases} \tag{1}$$

where $\mathbf{y}_t$, $\mathbf{y}_0 \in \mathbb{R}^N$, $\mathbf{\Phi_k} \in \mathbb{R}^{N \times N}$, and $\underline{T_k} = \{T_{k-1} + 1, T_{k-1} + 2, \dots, T_k\}$.

ASSUMPTIONS

1. The stochastic driver $\{\boldsymbol{\varepsilon}_t\}_{t \in \underline{T}}$ is i.i.d. vector white noise with mean zero and covariance matrix $\boldsymbol{\Omega}$.

2. The series is locally stationary, i.e. all eigenvalues of the $\mathbf{\Phi_k}$ are inside the unit circle for all k. The local means are $\boldsymbol{\mu}_k = (\mathbf{I} - \mathbf{\Phi}_k)^{-1}\mathbf{c}_k$.

3. The break fractions

$$\lambda_k := \frac{T_k - T_{k-1}}{T}$$

remain constant as the sample size T grows to infinity.

THEOREM I

VAR(1) is estimated globally on $\mathbf{y}_t$ and the parameter changes are ignored. Then, the least squares estimator satisfies in the probability limit

$$\hat{\mathbf{\Phi}}\Bigg[\sum_{i,j\in\underline{K},\, i\neq j}\lambda_i\lambda_j(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)' + \lim_{T\to\infty}\frac{2}{T}\sum_{k=1}^{K}\sum_{t=T_{k-1}+1}^{T_k}\sum_{j=1}^{t-T_{k-1}-1}\mathbf{\Phi}_k^{j-1}\mathbf{\Omega}(\mathbf{\Phi}_k^{j-1})'\Bigg]$$

$$= \sum_{i,j\in\underline{K},\, i\neq j}\lambda_i\lambda_j(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)' + \lim_{T\to\infty}\frac{2}{T}\sum_{k=1}^{K}\mathbf{\Phi}_k\sum_{t=T_{k-1}+1}^{T_k}\sum_{j=1}^{t-T_{k-1}-1}\mathbf{\Phi}_k^{j-1}\mathbf{\Omega}(\mathbf{\Phi}_k^{j-1})'$$

DISCUSSION

- In the scalar case, as $\sum_{i,j\in\underline{K},\, i\neq j} \lambda_i\lambda_j(\mu_i-\mu_j)^2$ grows large relative to σ^2, $\widehat{\phi} \to 1$. As σ^2 grows large relative to $\sum_{i,j\in\underline{K},\, i\neq j} \lambda_i\lambda_j(\mu_i-\mu_j)^2$,

$$\widehat{\phi} \to \frac{\phi_1\omega_1 + \phi_2\omega_2 + \ldots + \phi_K\omega_K}{\omega_1 + \omega_2 + \ldots + \omega_K}, \quad \omega_k = \frac{\lambda_k}{1-\phi_k^2}.$$

- AR(p) models can be written as VAR(1), so the Theorem applies. The result is that as $\sum_{i,j\in\underline{K},\, i\neq j} \lambda_i\lambda_j(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)'$ grows large,

$$\sum_{j=1}^{p} \widehat{\phi} \to 1.$$

  Similarly, for VAR(p), $\sum_{j=1}^{p} \hat{\boldsymbol{\Phi}}_j$ displays a unit eigenvalue.

- In the case of a single break in a single component of the VAR(1), the eigenvector corresponding to eigenvalue 1 is the square of the change-point detector proposed in Bai 1994, 1997.

OTHER ESTIMATORS

- Usual moment condition

$$\mathbb{E}_{t-1}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbb{E}_{t-1}\,\hat{\boldsymbol{\varepsilon}}_t = \mathbf{0},\ t \in \underline{T}$$

  cannot hold since

$$\mathbb{E}_{t-1}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0} = (\boldsymbol{\Phi}_k - \mathbb{E}_{t-1}\hat{\boldsymbol{\Phi}})\mathbf{y}_{t-1},\ t \in \underline{T_k},\ k \in \underline{K}.$$

- Instead: Consider estimators that satisfy asymptotically

$$\mathbb{E}_{T_{k-1}}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0},\ t \in \underline{T_k},\ k \in \underline{K}.$$

THEOREM II

VAR(1) is estimated globally on $\mathbf{y}_t$ and the parameter changes are ignored. Let the estimator $\hat{\mathbf{\Phi}}$ satisfy asymptotically

$$\mathbb{E}_{T_{k-1}}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0},\ t \in \underline{T_k},\ k \in \underline{K}.$$

Then,

$$\hat{\mathbf{c}} = \mathbf{0} \text{ and } \hat{\mathbf{\Phi}} = \mathbf{I}$$

PROOF 1

$$\mathbb{E}_{T_{k-1}}(\hat{\mathbf{y}}_t) = \mathbb{E}_{T_{k-1}}(\mathbf{y}_t),\ t \in \underline{T_k},\ k \in \underline{K}$$

$$= \mathbb{E}_{T_{k-1}}\left[\mathbf{\Phi}_k^{t-T_{k-1}}\mathbf{y}_{T_{k-1}} + (\mathbf{I}-\mathbf{\Phi}_k)^{-1}(\mathbf{I}-\mathbf{\Phi}^{t-T_{k-1}})\mathbf{c}_k + \sum_{j=0}^{t-T_{k-1}-1}\mathbf{\Phi}_k^j\boldsymbol{\varepsilon}_{t-j}\right]$$

$$\mathbb{E}_{T_{k-1}}(\hat{\mathbf{y}}_t) = \mathbb{E}_{T_{k-1}}(\mathbb{E}_{t-1}(\hat{\mathbf{y}}_t)) = \mathbb{E}_{T_{k-1}}(\hat{\mathbf{c}} + \hat{\mathbf{\Phi}}\mathbf{y}_{t-1})$$

$$= \hat{\mathbf{c}} + \hat{\mathbf{\Phi}}\mathbb{E}_{T_{k-1}}\left[\mathbf{\Phi}_k^{t-T_{k-1}-1}\mathbf{y}_{T_{k-1}} + (\mathbf{I}-\mathbf{\Phi}_k)^{-1}(\mathbf{I}-\mathbf{\Phi}_k^{t-T_{k-1}-1})\mathbf{c}_k + \sum_{j=0}^{t-T_{k-1}-2}\mathbf{\Phi}_k^j\boldsymbol{\varepsilon}_{t-j}\right].$$

$$\mathbf{\Phi}_k^{t-T_{k-1}}(\mathbf{y}_{T_{k-1}} - \boldsymbol{\mu}_k) + \boldsymbol{\mu}_k$$

$$= \hat{\mathbf{c}} + \hat{\mathbf{\Phi}}\left[\mathbf{\Phi}_k^{t-T_{k-1}-1}(\mathbf{y}_{T_{k-1}} - \boldsymbol{\mu}_k) + \boldsymbol{\mu}_k\right],\ t \in \underline{T_k},\ k \in \underline{K}.$$

PROOF 2

As $T \to \infty$,

$$\boldsymbol{\mu}_k = \hat{\mathbf{c}} + \hat{\boldsymbol{\Phi}} \boldsymbol{\mu}_k, \; k \in \underline{K}$$

$$\Longrightarrow \quad \hat{\mathbf{c}} = \mathbf{0}, \; \hat{\boldsymbol{\Phi}} = \mathbf{I}.$$

Q.E.D.

DISCUSSION

- Result holds as $T \to \infty$. *Does not depend on break size.*

- LS estimator of VAR does not satisfy $\mathbb{E}_{T_{k-1}}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0},\ t \in \underline{T_k},\ k \in \underline{K}$ but only

$$\mathbb{E}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0},\ t \in \underline{T_k},\ k \in \underline{K}.$$

- MLE of GARCH does satisfy the condition.

$$\begin{aligned} r_t &= \varepsilon_t,\ t \in \underline{T} \\ \varepsilon_t | \mathcal{F}_{t-1} &\sim \mathcal{N}(0, h_t), \\ h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \end{aligned}$$

ARGUMENT FOR GARCH

Lumsdaine (1996):

$$\mathbb{E}(L(\hat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta})) = \mathbb{E}\left[\log \frac{h_t}{\widehat{h}_t} - \frac{h_t}{\widehat{h}_t}\right] \leq \mathbb{E}f(1),$$

$f(x) = \log x - x$ attains maximum at $x = 1$: $\mathbb{E}(L(\hat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}))$ attains maximum at $x = h_t/\widehat{h}_t = 1$ or

$$h_t = \widehat{h}_t \text{ but } h_t = \mathbb{E}_{t-1}h_t = \mathbb{E}_{t-1}\widehat{h}_t.$$

Therefore, optimality at $\mathbb{E}_{t-1}(h_t - \widehat{h}_t) = 0.$

Doesn't work in the case of breaks! Only way to satisfy $h_t \approx \widehat{h}_t$ asymptotically

$$\mathbb{E}_{T_{k-1}}(h_t - \widehat{h}_t) = 0.$$

Equivalent to:

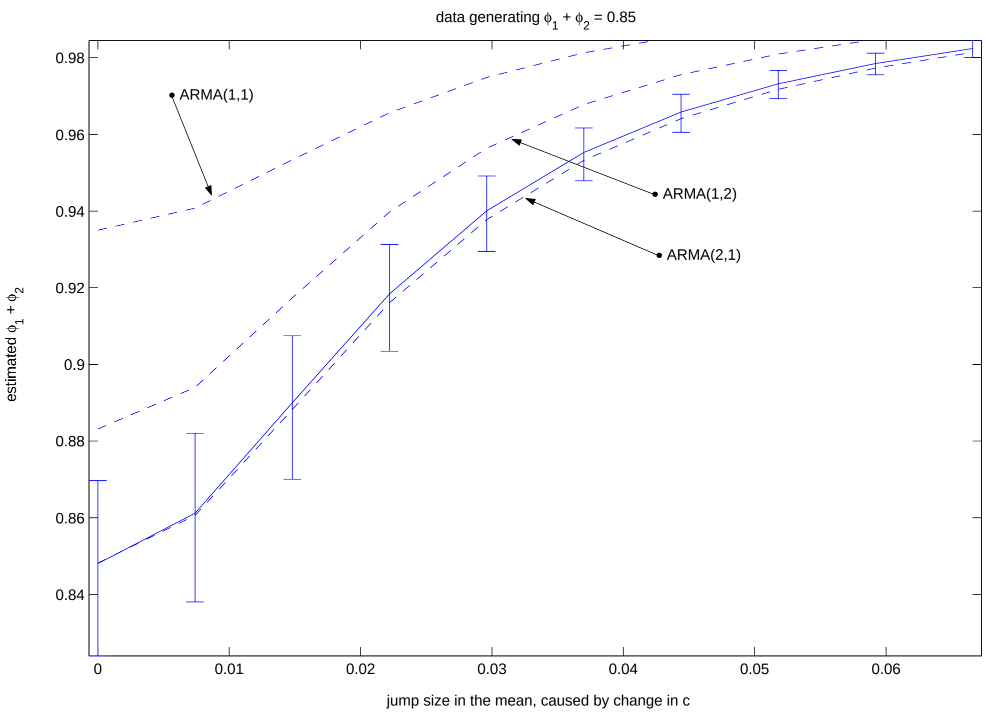
$$\mu_k = \widehat{\alpha}_0 + \left(\sum_{j=1}^{q} \widehat{\alpha}_j + \sum_{j=1}^{p} \widehat{\beta}_j\right)\mu_k,\ k \in \underline{K}.$$

Set $\widehat{\alpha}_0 = 0$ and $\sum_{j=1}^{q} \widehat{\alpha}_j + \sum_{j=1}^{p} \widehat{\beta}_j = 1.$

SIMULATIONS: ARMA

DGP: ARMA(2,2), $\phi_1 = 0.20$, $\phi_2 = 0.65$, 10 different break sizes in c at $T/2$.

Plot of $\widehat{\phi}_1 + \widehat{\phi}_2$:

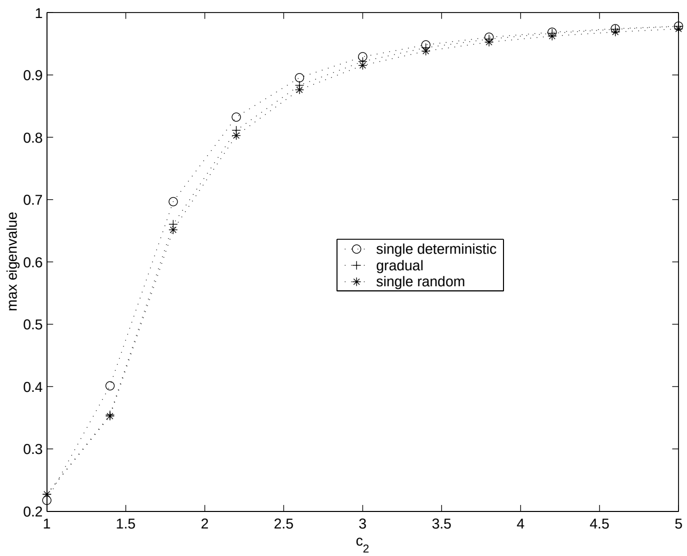


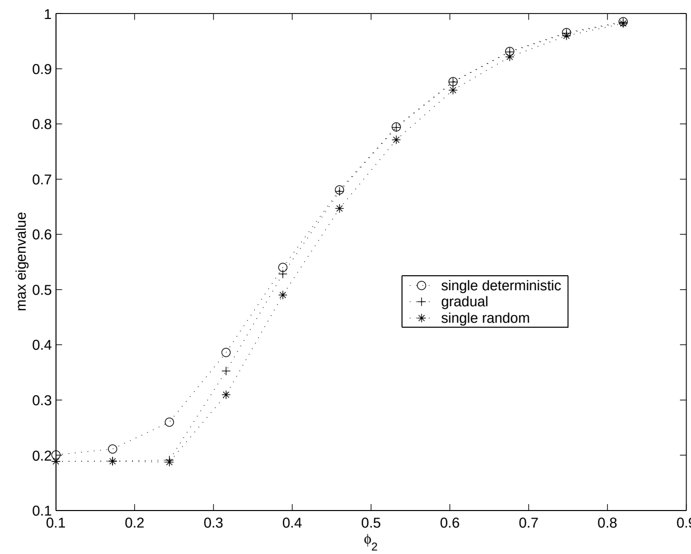
SIMULATIONS: VAR

DGP: VAR(1), 11 different break sizes at $T/2$, single, gradual, and random breaks.

Largest eigenvalue in $\hat{\mathbf{\Phi}}$ for

changes in c, changes in $\mathbf{\Phi}$.

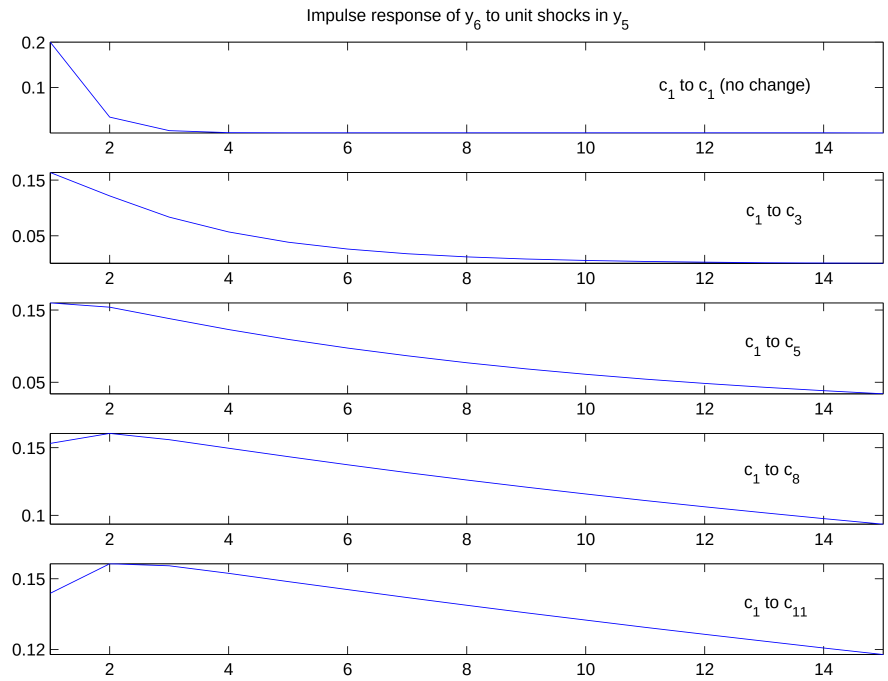




SIMULATIONS: IRF

DGP: VAR(1), 5 different single break sizes at $T/2$. The 6th entry in the VAR has lag 1 of the 5th entry in the DGP. Lag coefficient is 0.20.

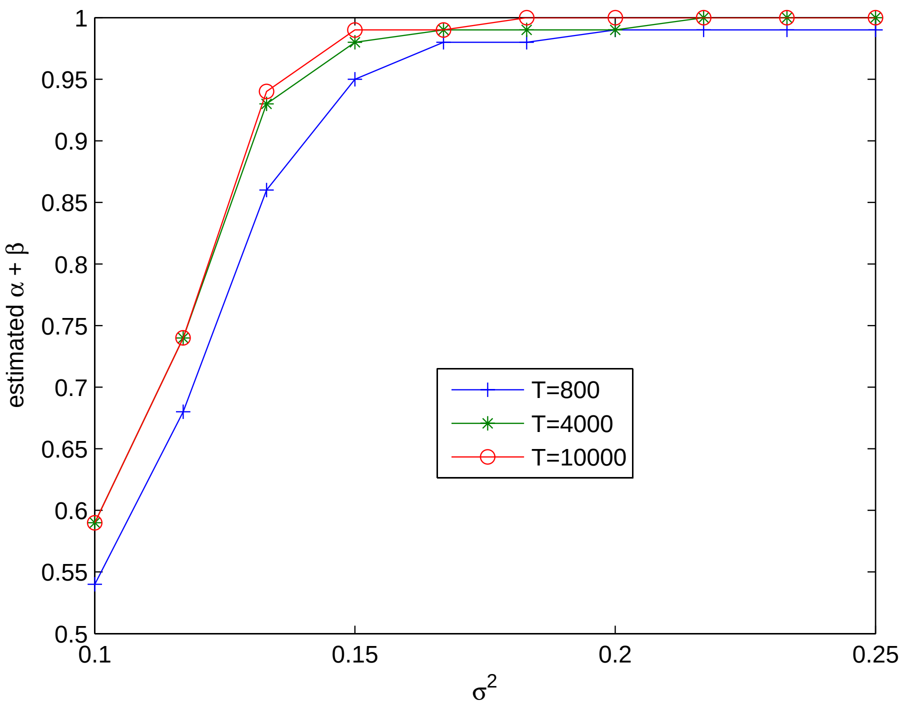
Impulse response function of $\mathbf{y}_{6,t}$ to unit shocks in $\mathbf{y}_{5,t}$.



SIMULATIONS: GARCH

DGP: GARCH(1,1), $\alpha_0 = 1.65\text{e-}5$, $\alpha = 0.10$, 10 different break sizes in β at $T/2$.(H. 2005)

Plot of $\widehat{\alpha} + \widehat{\beta}$:



CONCLUSIONS

- Neglected parameter changes lead to overestimation of persistence in autoregressive model models.

- Impulse response functions are overestimated.

- AR(1): $\widehat{\phi} \to 1$; AR(p): $\sum \widehat{\phi}_j \to 1$; VAR(1): $\hat{\mathbf{\Phi}}$ has unit eigenvalue, VAR(p): $\sum \hat{\mathbf{\Phi}}_j$ has unit eigenvalue.

- The least squares estimator shows trade-off between break size and variance.

- Estimators that satisfy $\mathbb{E}_{T_{k-1}}(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0},\ t \in \underline{T_k},\ k \in \underline{K}$ grow to one as $T \to \infty$.