Forecasting VARMA processes: VAR models vs. subspace-based state space models

Segismundo Izquierdo Millán

Cesáreo Hernández Iglesias

Juan del Hoyo Bernat

Cyprus, June 2006

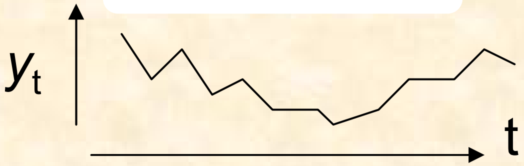
CEF2006

Index

1 Introduction. Objective

Observations

$$(\mathbf{I}_m + \mathbf{A}_1 L + \ldots + \mathbf{A}_p L^p) \ \boldsymbol{y}_t = (\mathbf{I}_m + \mathbf{B}_1 L + \ldots + \mathbf{B}_q L^q) \ \boldsymbol{e}_t$$



System identification method

Forecasting

Models

Vector Auto-Regressive (VAR)

$$(\mathbf{I}_m + \mathbf{\Phi}_1 L + \ldots + \mathbf{\Phi}_p L^p) \ \boldsymbol{y}_t = \boldsymbol{e}_t$$

State Space (SS)

$$\boldsymbol{z}_{t+1} = \mathbf{A} \, \boldsymbol{z}_t + \mathbf{K} \, \boldsymbol{e}_t \qquad \text{State transition equation}$$

$$\boldsymbol{y}_t = \mathbf{C} \, \boldsymbol{z}_t + \boldsymbol{e}_t \qquad \text{Observation equation}$$

Motivation

VAR	vs.	VARMA
•One-step ML estimation (Least Squares)		•Involved ML estimation (iterative)
•Simple specification		•Difficult specification
•“Not as general” (finite order)		•Parsimonious

State Space with subspace methods

- One-step estimation (can be “refined”)
- The simplest specification
- Equivalent to VARMA

Methodology: Subspace Methods

Orthogonal-projection-based linear forecast $\approx$ **State-based linear forecast**

$$\underbrace{\left[\begin{array}{c:c:c} y_{p+1}^{f}/y_{p}^{p} & y_{p+1}^{f}/y_{p+1}^{p} & \cdots \end{array}\right]}_{U\,S\,V} \approx \underbrace{\begin{bmatrix} C \\ C\,A \\ \cdots \\ C\,A^{f-1} \end{bmatrix}}_{O_f} \underbrace{\left[\begin{array}{c:c:c} z_{p+1|p} & z_{p+2|p+1} & \cdots \end{array}\right]}_{Z}$$

$$\begin{tikzcd} & U\,S\,V \arrow[dl] \arrow[d] \arrow[dr] & \\ \hat{O}_f & \hat{n} & \hat{Z} \end{tikzcd}$$

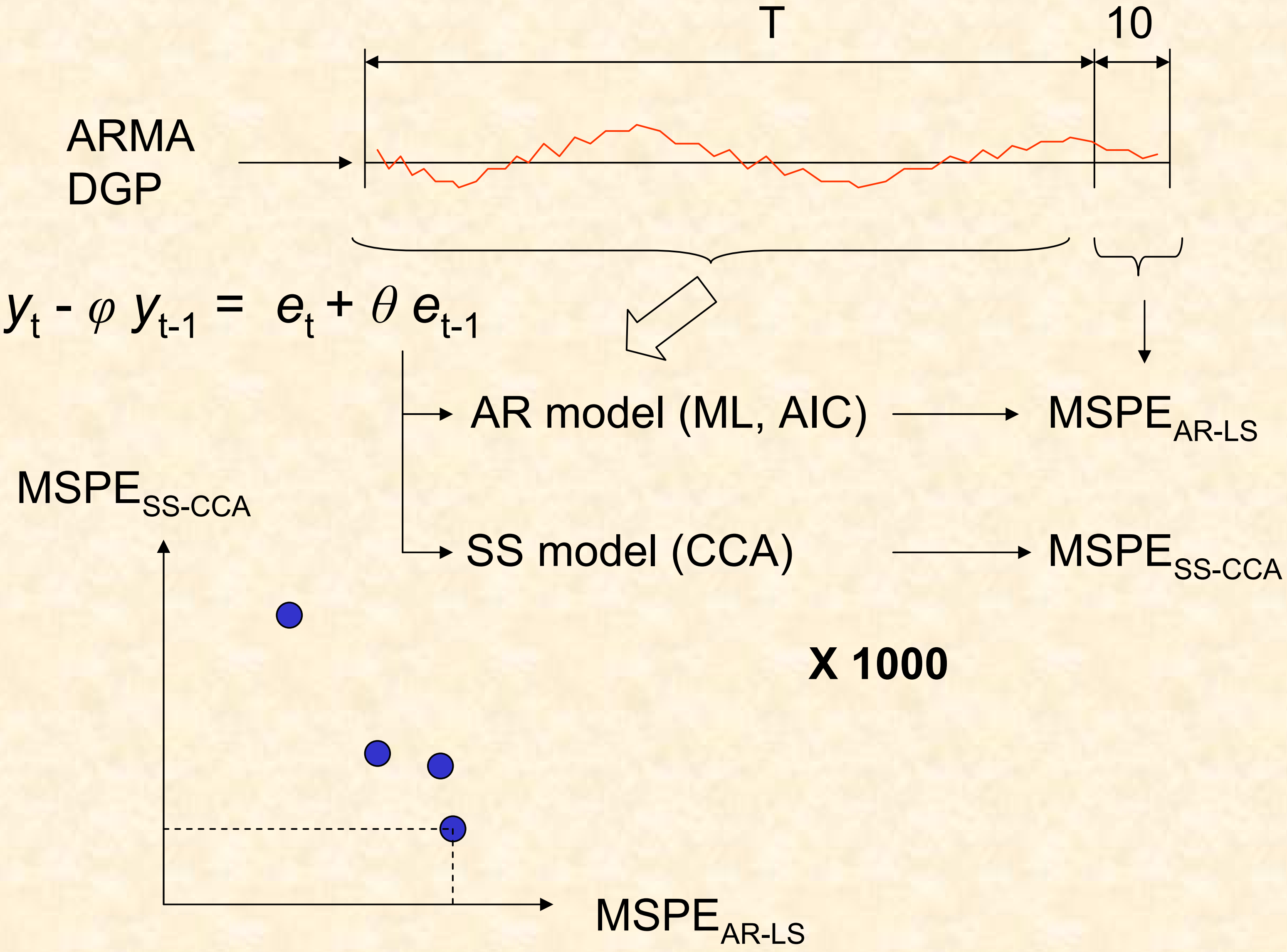
$$\Longrightarrow \begin{aligned} z_{t+1} &= A\, z_t + K\, e_t \\ y_t &= C\, z_t + e_t \end{aligned}$$

Consistent

CCA -> ML

Index

2.1 ARMA(1,1): Forecasting



ARMA DGP

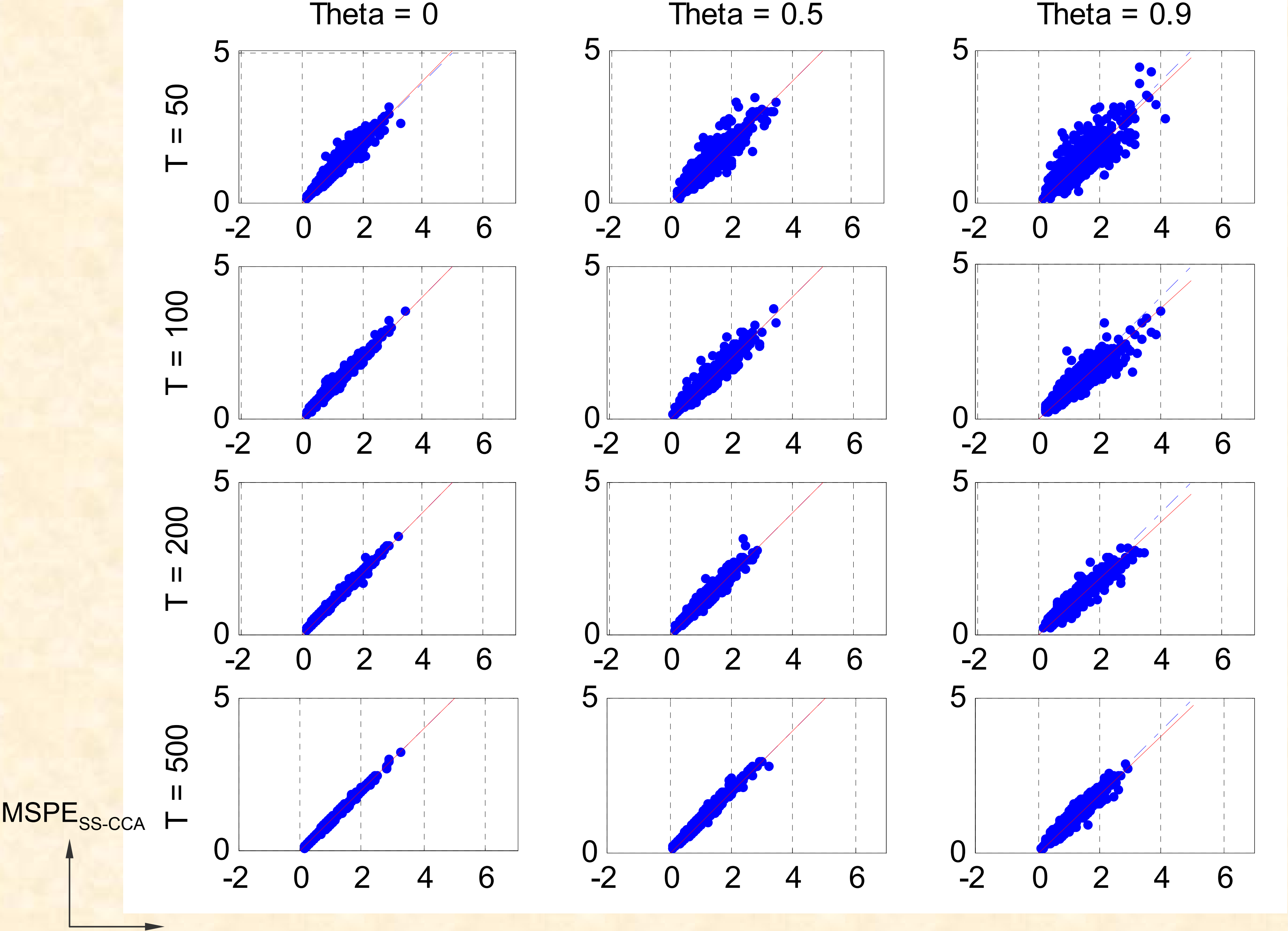
$$y_t - \varphi\, y_{t-1} = e_t + \theta\, e_{t-1}$$

AR model (ML, AIC) → MSPE$_{AR\text{-}LS}$

SS model (CCA) → MSPE$_{SS\text{-}CCA}$

X 1000

Red: regression line



1000-replicate series of DGP1 with $\varphi = .9$

$$y_t - \varphi\, y_{t-1} = e_t + \theta\, e_{t-1}$$

2.1 ARMA(1,1): Forecasting

	θ		
T	*0*	*0.5*	*0.9*
50	1.02	1.01	0.94
100	1.01	0.99	0.90
200	1.00	1.00	0.92
500	1.00	1.00	0.95

Table 2. Slopes of the regressions of $\text{MSPE}_{\text{SS-CCA}}$ on $\text{MSPE}_{\text{AR-LS}}$ calculated in a 1000-replicate series of DGP1 with $\varphi = .9$

2.1 ARMA(1,1): Forecasting

	θ		
T	0	0.5	0.9
50	**370.5**	**472**	**575**
100	**420.5**	514	**653**
200	**447.5**	**544.5**	**677**
500	**421**	512	**644**

Binomial test

$$P(x \text{ out of } N) = \binom{N}{x} (\tfrac{1}{2})^{x} \, (1 - \tfrac{1}{2})^{N-x}$$

$$\mathbf{P(459 \leq x \leq 541) = 0.991}$$

Table 3. Number of samples in which $MSPE_{SS\text{-}CCA} < MSPE_{AR\text{-}LS}$, plus half the number of samples in which $MSPE_{SS\text{-}CCA} = MSPE_{AR\text{-}LS}$, out of a 1000-replicate series. Significant values for a binomial test (H_0 : $P(MSPE_{SS\text{-}CCA} < MSPE_{AR\text{-}LS}) = P(MSPE_{AR\text{-}LS} < MSPE_{SS\text{-}CCA})$; $\alpha < 0.01$, two-sided test) are shown in bold.

2.1 ARMA(1,1): Forecasting

$$\frac{\mathrm{MSPE}_{\mathrm{SS\text{-}CCA}} - \mathrm{MSPE}_{\mathrm{AR\text{-}LS}}}{\mathrm{MSPE}_{\mathrm{AR\text{-}LS}}}$$

$$d_{\mathrm{i}} = (\mathrm{MSPE}_{\mathrm{SS\text{-}CCA}} - \mathrm{MSPE}_{\mathrm{AR\text{-}LS}})_{\mathrm{i}}$$

$$S = \sqrt{N}\,\frac{\bar{d}_i}{std(d_i)}$$

	θ		
T	*0*	*0.5*	*0.9*
50	**2%**	**3%**	**-3%**
100	**1%**	0%	**-7%**
200	0%	0%	**-7%**
500	0%	0%	**-4%**

	θ		
T	*0*	*0.5*	*0.9*
50	**6,01**	**4,31**	**-3,79**
100	**4,95**	0,55	**-12,00**
200	1,87	0,35	**-13,24**
500	2,60	-1,00	**-10,35**

Table 4. Left: total increase in $\mathrm{MSPE}_{\mathrm{SS\text{-}CCA}}$ with respect to $\mathrm{MSPE}_{\mathrm{AR\text{-}LS}}$ out of a 1000-replicate series of DGP1 with $\varphi = .9$. Right: corresponding values of the statistic *S*. Significant values ($\alpha = 0.01$) are shown in bold.

2.1 ARMA(1,1): Forecasting

$\text{MSPE}_{\text{ARMA11-ML}}$ vs. $\text{MSPE}_{\text{AR-LS}}$

	θ		
T	*0*	*0.5*	*0.9*
50	**1%**	**-5%**	**-17%**
100	0%	**-3%**	**-13%**
200	0%	**-2%**	**-9%**
500	0%	**-1%**	**-5%**

$\text{MSPE}_{\text{ARMA11-ML}}$ vs. $\text{MSPE}_{\text{SS-CCA}}$

	θ		
T	*0*	*0.5*	*0.9*
50	-1%	**-7%**	**-14%**
100	**-1%**	**-3%**	**-6%**
200	0%	**-2%**	**-2%**
500	0%	**-1%**	**-1%**

$\text{MSPE}_{\text{SS-ML}}$ vs. $\text{MSPE}_{\text{SS-CCA}}$

	θ		
T	*0*	*0.5*	*0.9*
50	0%	**-3%**	**-4%**
100	-1%	**-2%**	**-2%**
200	0%	**-2%**	-1%
500	0%	**-1%**	-1%

$\text{MSPE}_{\text{ARMA11-ML}}$ vs. $\text{MSPE}_{\text{SS1-ML}}$

	θ		
T	*0*	*0.5*	*0.9*
50	0%	1%	1%
100	0%	0%	0%
200	0%	0%	0%
500	0%	0%	0%

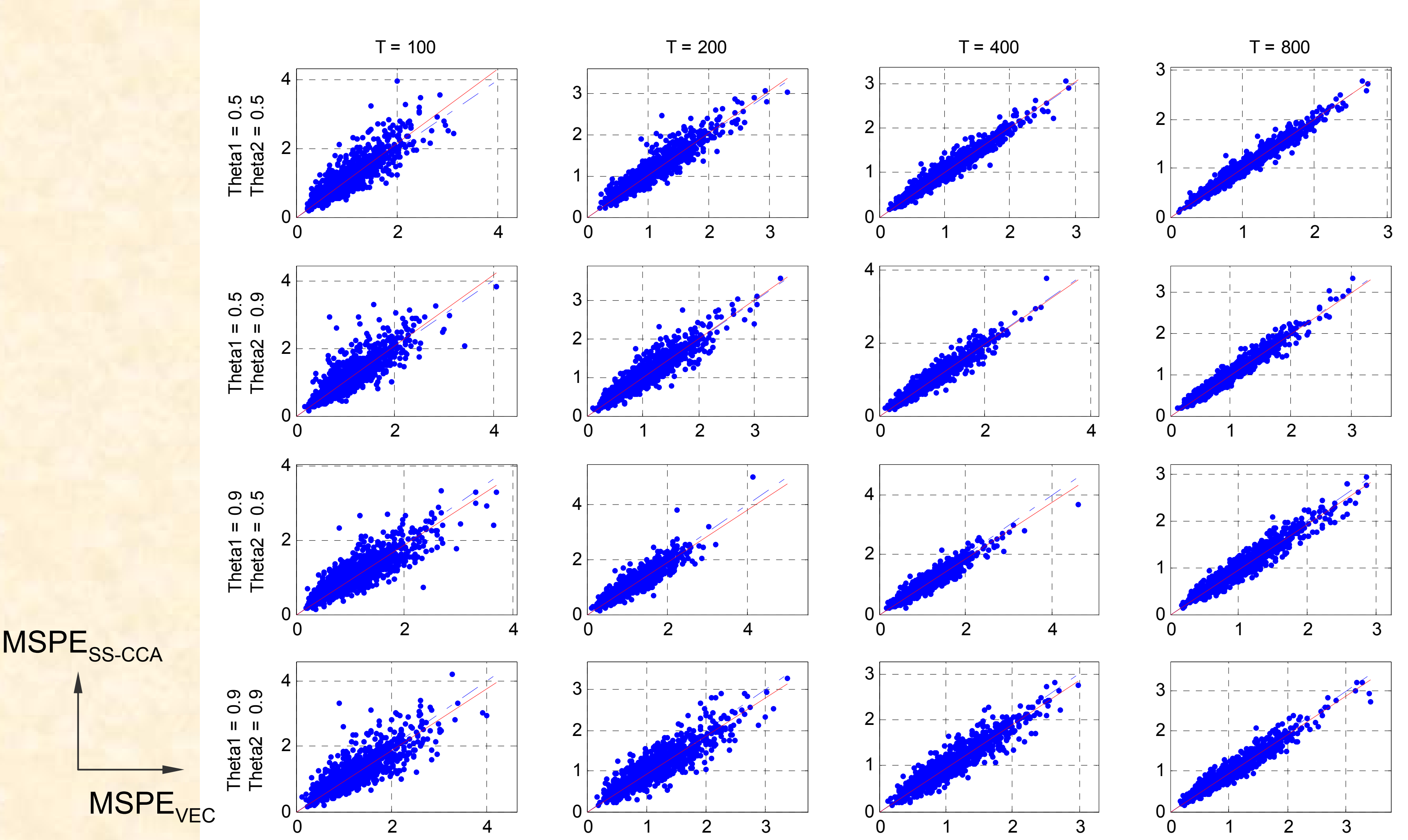
2.2 cointegrated VARMA: Forecasting

$$\Delta y_{1,t} = (1 + \theta_1 L)\, e_{1,t}$$

$$y_{2,t} = \gamma + \beta y_{1,t} + (1 + \theta_2 L)\, e_{2,t}$$

Johansen's VEC models

SS CCA models



2.2 Cointegrated VARMA: Forecasting

		y_1				y_2			
θ_1	θ_2	*T=100*	*200*	*400*	*800*	*T=100*	*200*	*400*	*800*
0.5	*0.5*	**10%**	**4%**	**2%**	1%	**11%**	**4%**	**2%**	1%
	0.9	**6%**	**3%**	0%	0%	-1%	-1%	**-2%**	**-1%**
0.9	*0.5*	**-3%**	**-4%**	**-5%**	**-3%**	**3%**	0%	**-2%**	**-2%**
	0.9	-2%	**-5%**	**-4%**	**-3%**	-2%	**-5%**	**-4%**	**-3%**

Table. Total increase in $\mathrm{MSPE}_{\mathrm{CCA}}$ with respect to $\mathrm{MSPE}_{\mathrm{VEC}}$ out of a 1000-replicate series of DGP2 with $\beta = 1$, $\sigma = 1$ and $\rho = .8$.

Index

3. Conclusions

Do SS-CCA models provide better forecasts than VAR models when forecasting processes with large MA components in their VARMA representation?

Our simulations indicate:

Yes, in the univariate case

Not an easy answer in the multivariate case (several MA components and several series to forecast), though in general large MA components seem to favour the SS models compared with the VAR models

But small sample sizes (≈ 50 univariate, 100 bivariate) damage the performance of subspace algorithms more than the performance of VAR (LS) models

Subspace SS models are quick and easy to obtain, and different to VARs-> Good complements or alternatives

Forecasting VARMA processes: VAR models vs. subspace-based state space models

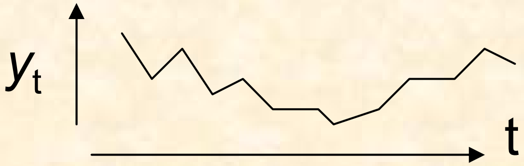
Segismundo Izquierdo Millán
Cesáreo Hernández Iglesias
Juan del Hoyo Bernat

Cyprus, June 2006

CEF2006

Index

Methodology: PEM



$$\longrightarrow M(\boldsymbol{\theta}) \longrightarrow \hat{\boldsymbol{y}}_t(\boldsymbol{\theta})$$

$$\big\downarrow$$

$$\varepsilon_t(\boldsymbol{\theta}) = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t(\boldsymbol{\theta})$$

$\mathbf{Y}_T$

$\boldsymbol{\theta}$

$$V_T(\boldsymbol{\theta}, \mathbf{Y}_T) = \frac{1}{T}\sum_{t=1}^{T} f(\varepsilon_t(\theta))$$

(Loss function for $\boldsymbol{\theta}$)

Quadratic criterion for $f(\cdot)$ $\longrightarrow$ Least Squares Estimates

Quadratic criterion for $f(\cdot)$
plus Gaussian zero-mean white noise innovations $\longrightarrow$ ML Estimates

Methodology: Subspace Methods

Orthogonal projection based linear forecast ≈ **State-based linear forecast**

$$\mathbf{y}_\mathbf{t}^{f} / \mathbf{y}_{\mathbf{t-1}}^{p}$$

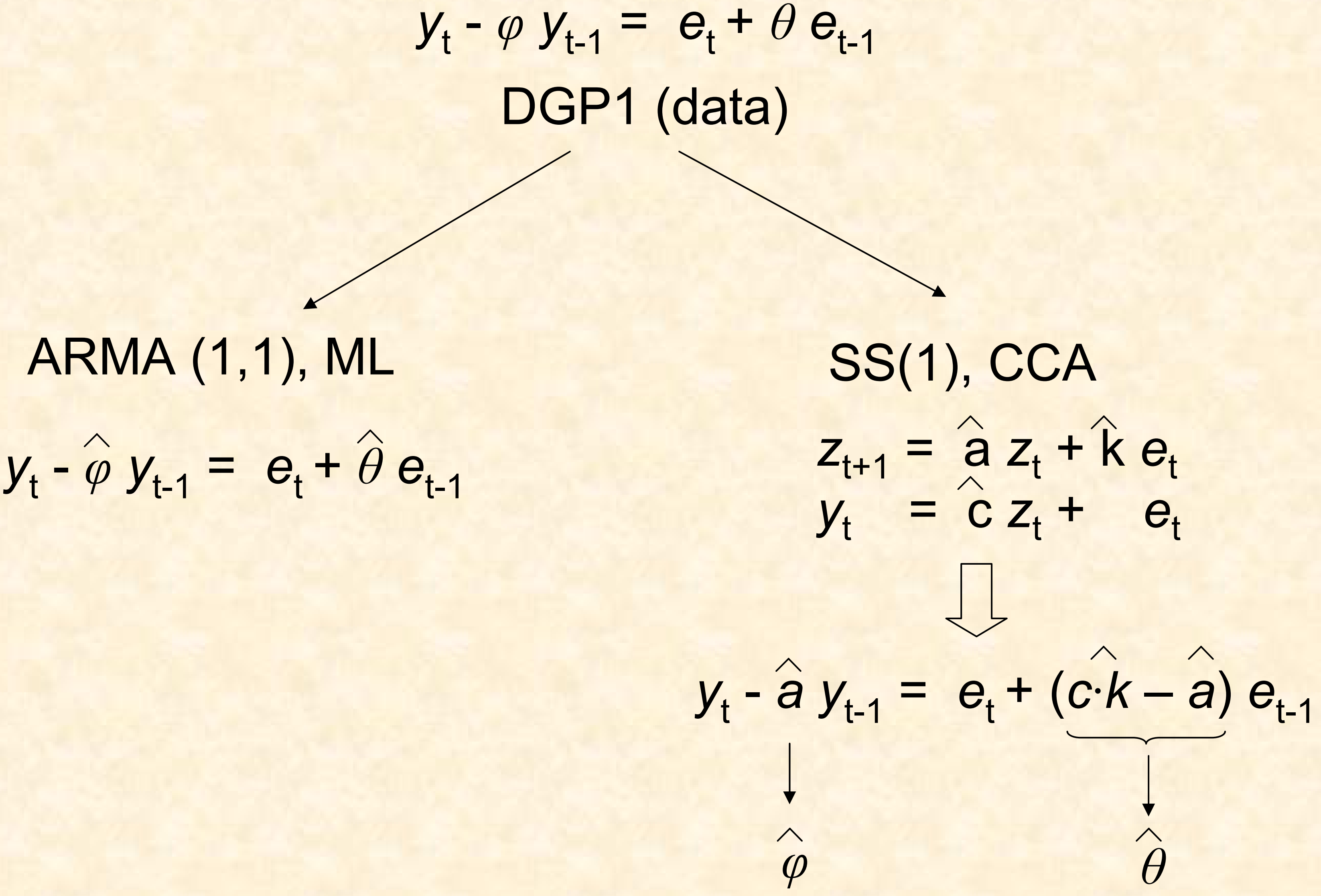
$$\mathbf{y}_\mathbf{t}^{f} = \begin{pmatrix} \mathbf{y}_\mathbf{t} \\ \mathbf{y}_\mathbf{t+1} \\ \cdots \\ \mathbf{y}_\mathbf{t+f-1} \end{pmatrix} \qquad \mathbf{y}_{\mathbf{t-1}}^{p} = \begin{pmatrix} \mathbf{y}_\mathbf{t-1} \\ \mathbf{y}_\mathbf{t-2} \\ \cdots \\ \mathbf{y}_\mathbf{t-p} \end{pmatrix}$$

$$\underbrace{E(\mathbf{y}_\mathbf{t}^{f}\, \mathbf{y}_{\mathbf{t-1}}^{p})E(\mathbf{y}_{\mathbf{t-1}}^{p}\, \mathbf{y}_{\mathbf{t-1}}^{p})^{-1}}_{\text{Estimated from the data}}\ \mathbf{y}_{\mathbf{t-1}}^{p}$$

$$\underbrace{\begin{pmatrix} \mathbf{C} \\ \mathbf{C\,A} \\ \cdots \\ \mathbf{C\,A}^{f-1} \end{pmatrix}}_{\mathbf{O}_f} \mathbf{z}_{t|t-1}$$

Index

3.1 ARMA(1,1): Recovering the system parameters



$$y_t - \varphi\, y_{t-1} = e_t + \theta\, e_{t-1}$$

DGP1 (data)

ARMA (1,1), ML

$$y_t - \hat{\varphi}\, y_{t-1} = e_t + \hat{\theta}\, e_{t-1}$$

SS(1), CCA

$$\begin{aligned} z_{t+1} &= \hat{a}\, z_t + \hat{k}\, e_t \\ y_t &= \hat{c}\, z_t + e_t \end{aligned}$$

$$\Downarrow$$

$$y_t - \underset{\hat{\varphi}}{\underset{\downarrow}{\hat{a}}}\, y_{t-1} = e_t + \underbrace{(\widehat{c\cdot k} - \hat{a})}_{\underset{\hat{\theta}}{\downarrow}}\, e_{t-1}$$

3.1 ARMA(1,1): Recovering the system parameters

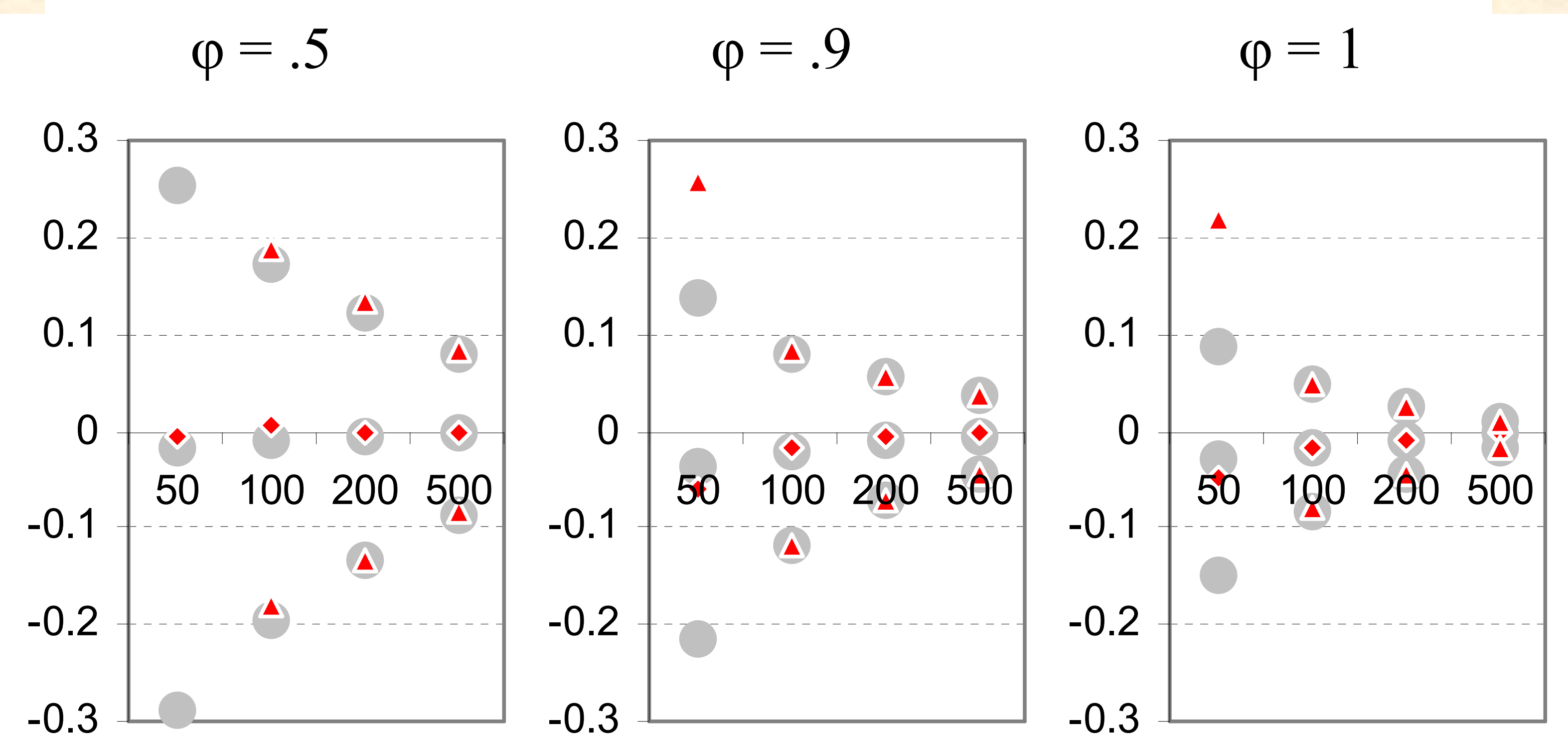
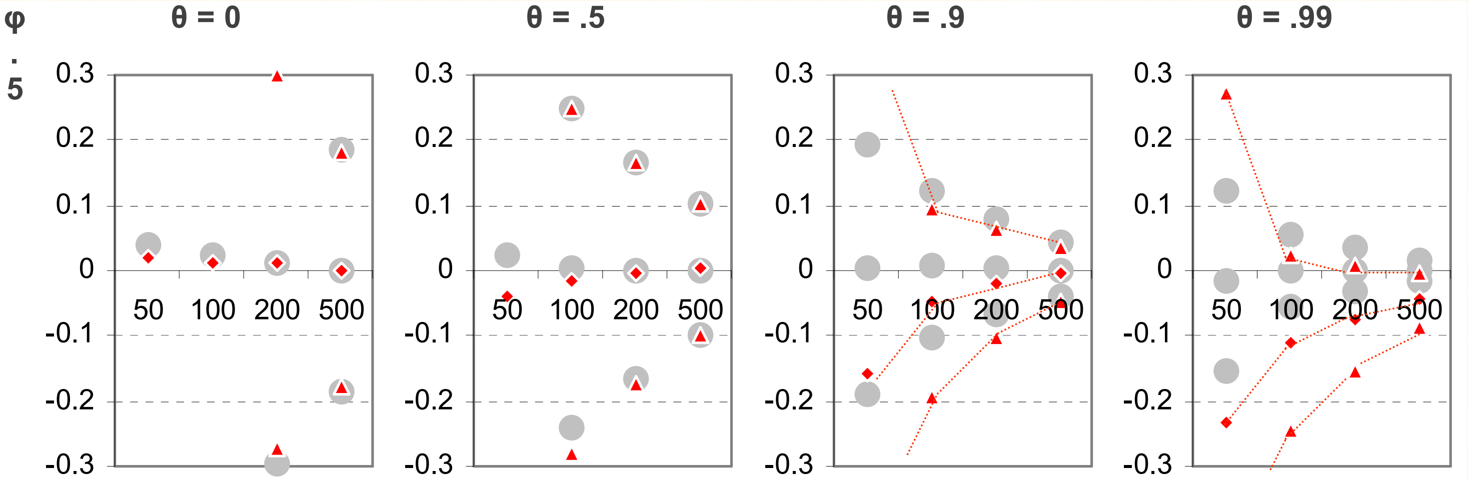


Table 1. Average bias and standard error for the estimates of φ according to the ARMA(1,1)_ML and SS(1)_CCA procedures, calculated in a 1000-replicate series of DGP1 with θ = .9

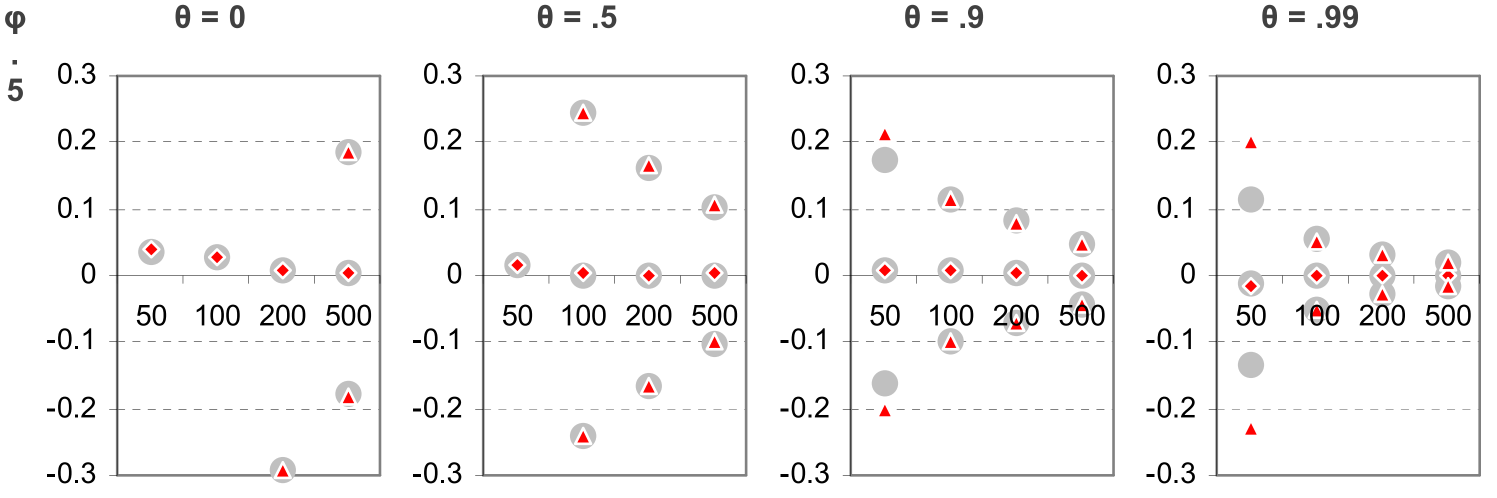
3.1 ARMA(1,1): Recovering the system parameters

Recovering θ : SS(1)_CCA vs. ARMA(1,1)_ML



Average biases +/- 2 s.e.

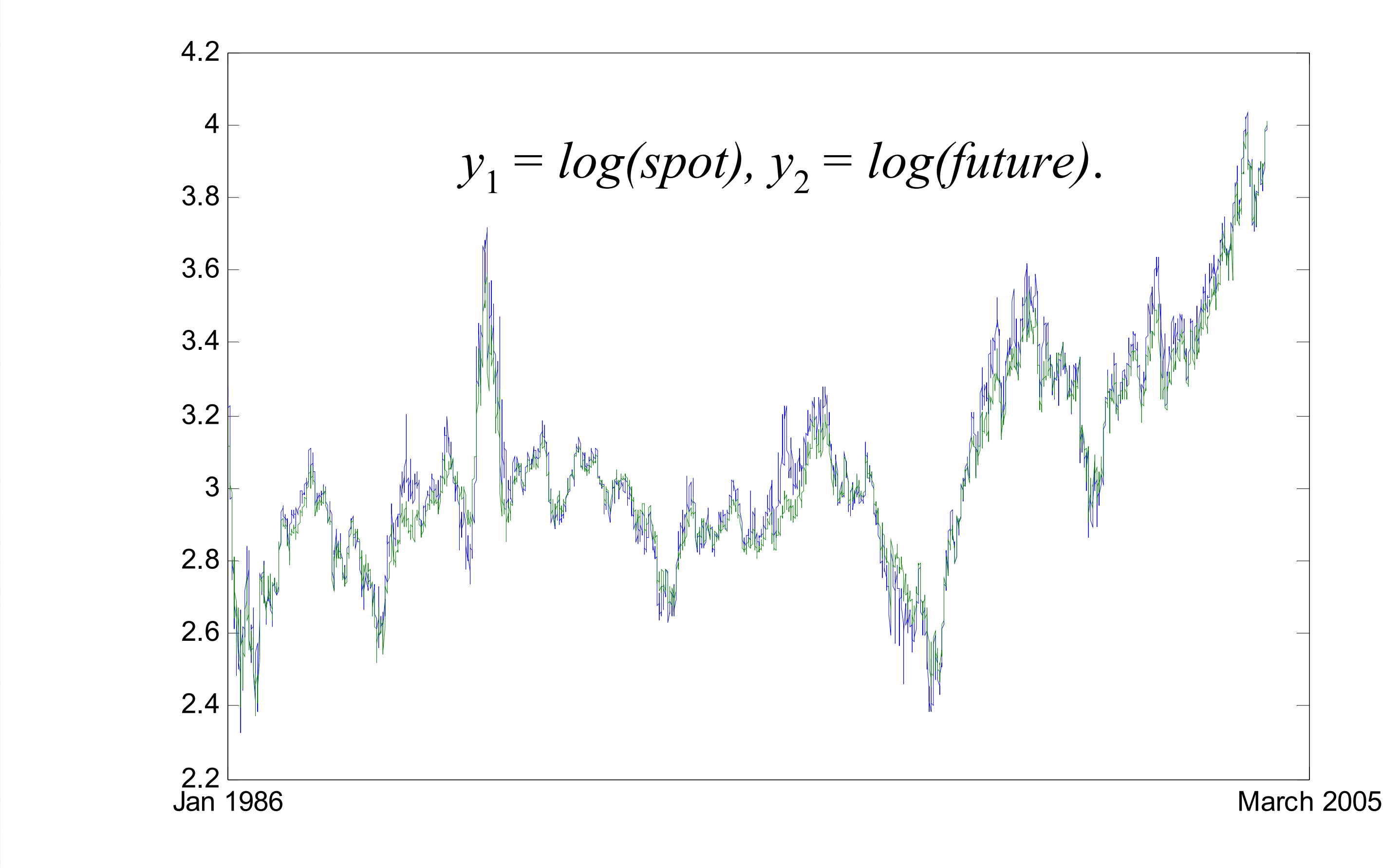
Recovering θ : SS(1)_ML vs. ARMA(1,1)_ML



Index

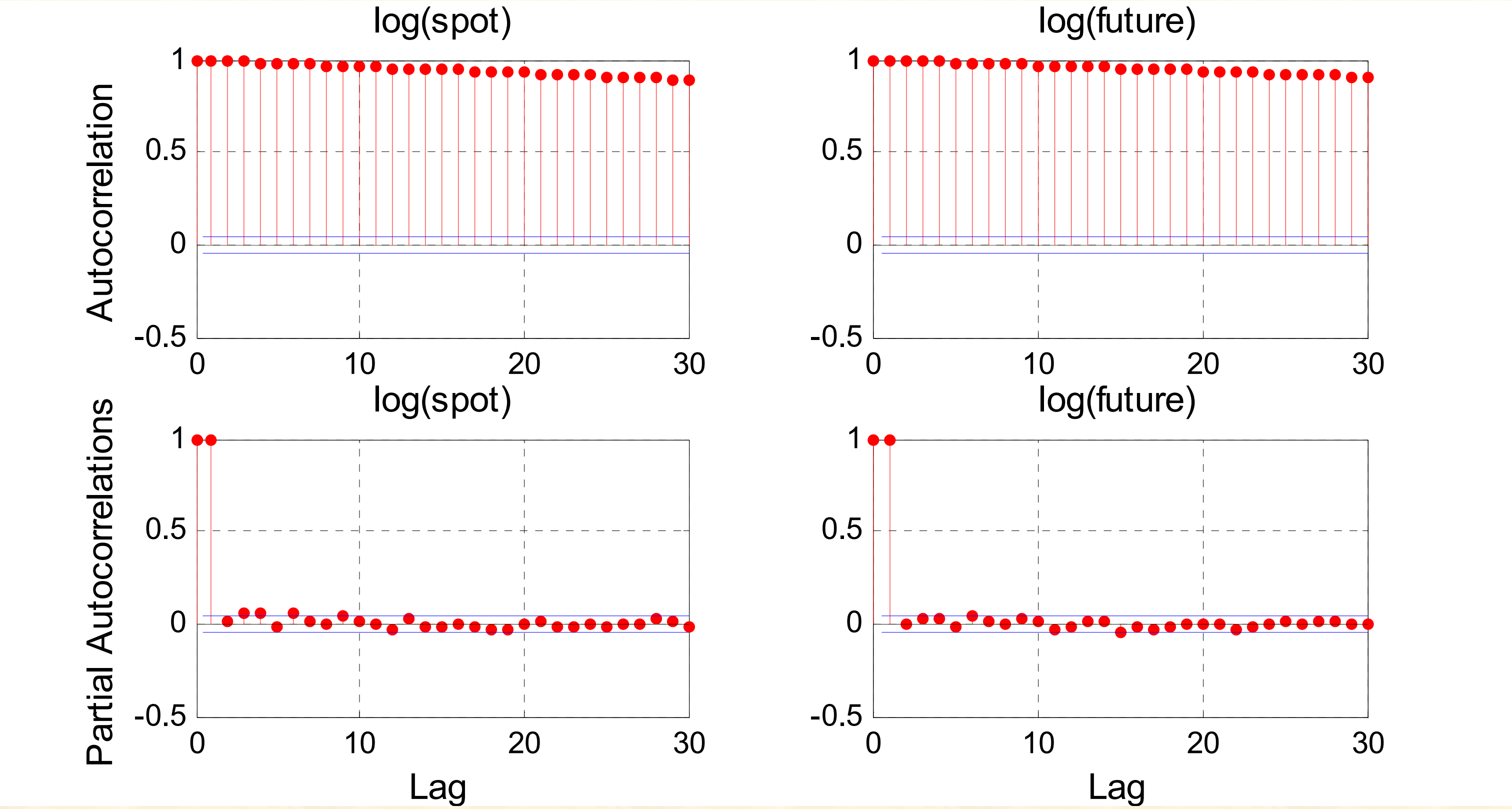
4. Practical case

4,805 crude oil daily prices between January 1986 and March 2005, as well as the corresponding “oil future contract” prices



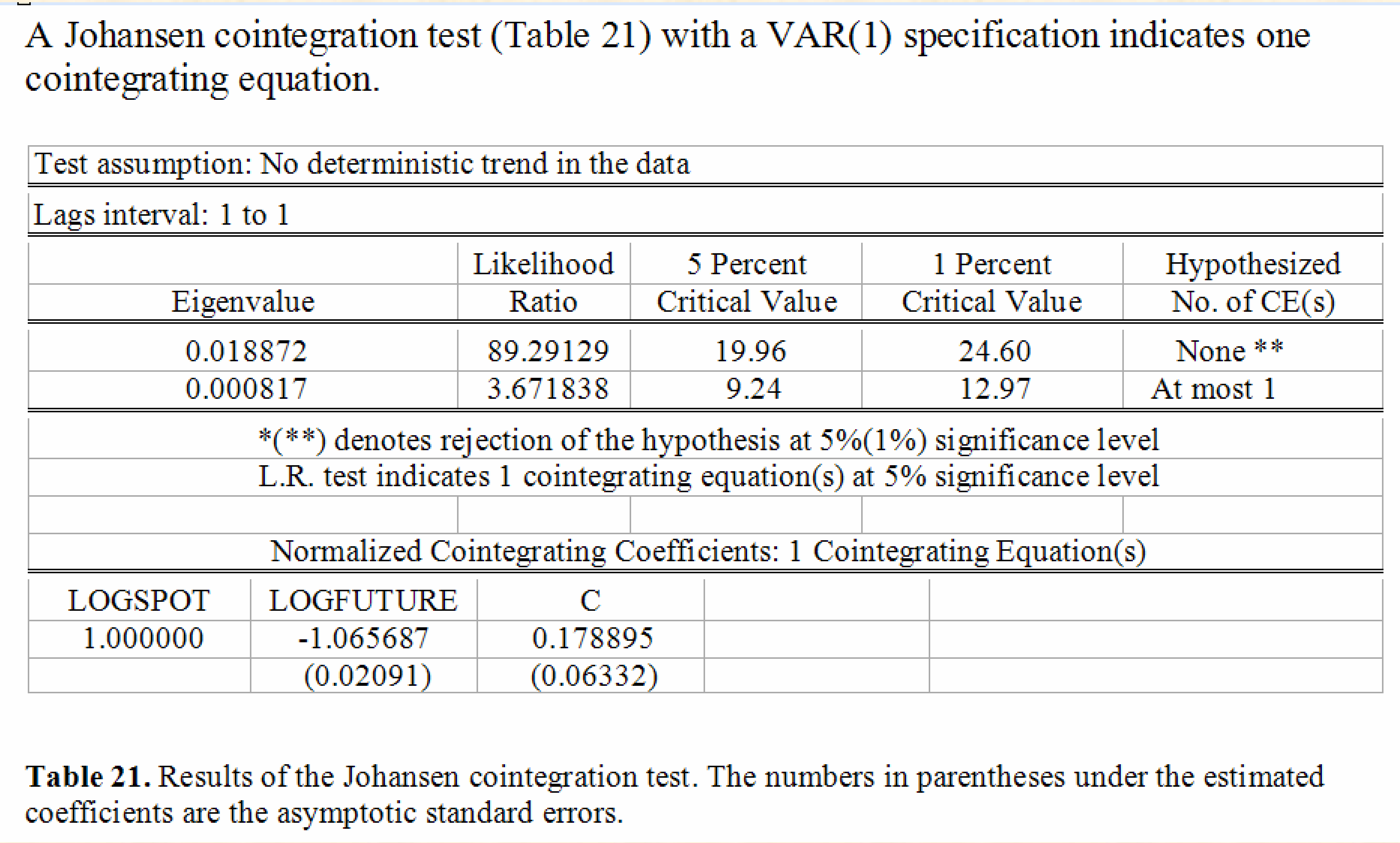
$$y_1 = log(spot), y_2 = log(future).$$

4. Practical case



For both series, if we assume no deterministic trend, the augmented Dickey-Fuller test with 1 lag (this number of lags was selected to eliminate correlation in the residuals), as well as the Philips-Perron test, do not reject the null of a unit root at the 5% level

4. Practical case



A Johansen cointegration test (Table 21) with a VAR(1) specification indicates one cointegrating equation.

Test assumption: No deterministic trend in the data

Lags interval: 1 to 1

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.018872	89.29129	19.96	24.60	None **
0.000817	3.671838	9.24	12.97	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

L.R. test indicates 1 cointegrating equation(s) at 5% significance level

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

LOGSPOT	LOGFUTURE	C		
1.000000	-1.065687	0.178895		
	(0.02091)	(0.06332)		

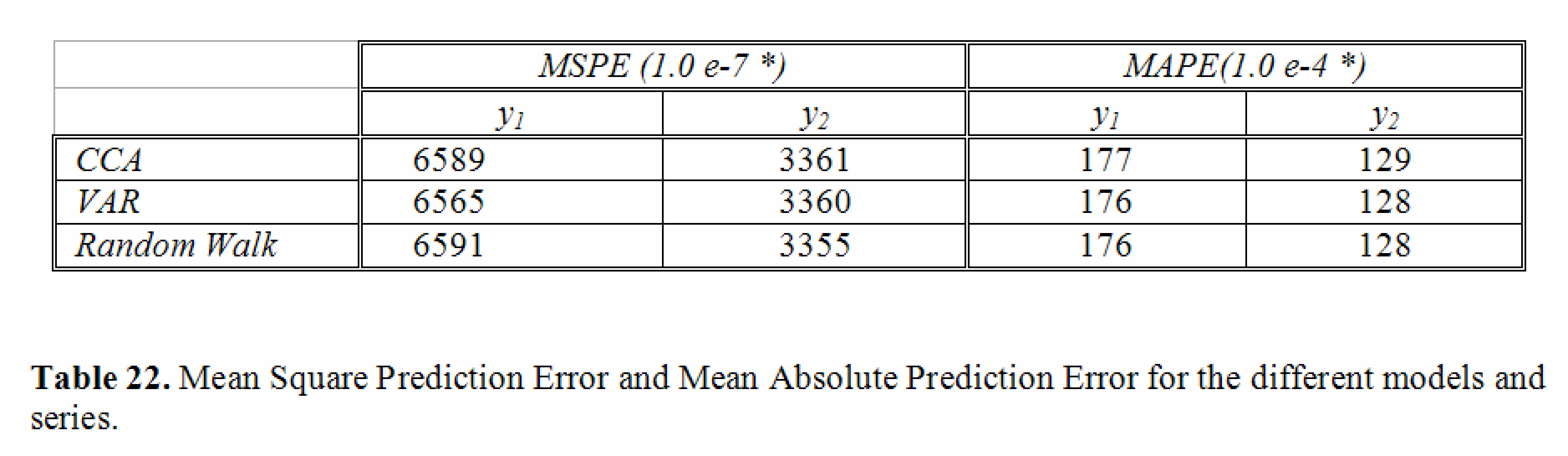
Table 21. Results of the Johansen cointegration test. The numbers in parentheses under the estimated coefficients are the asymptotic standard errors.

4. Practical case

One-step-ahead forecasts

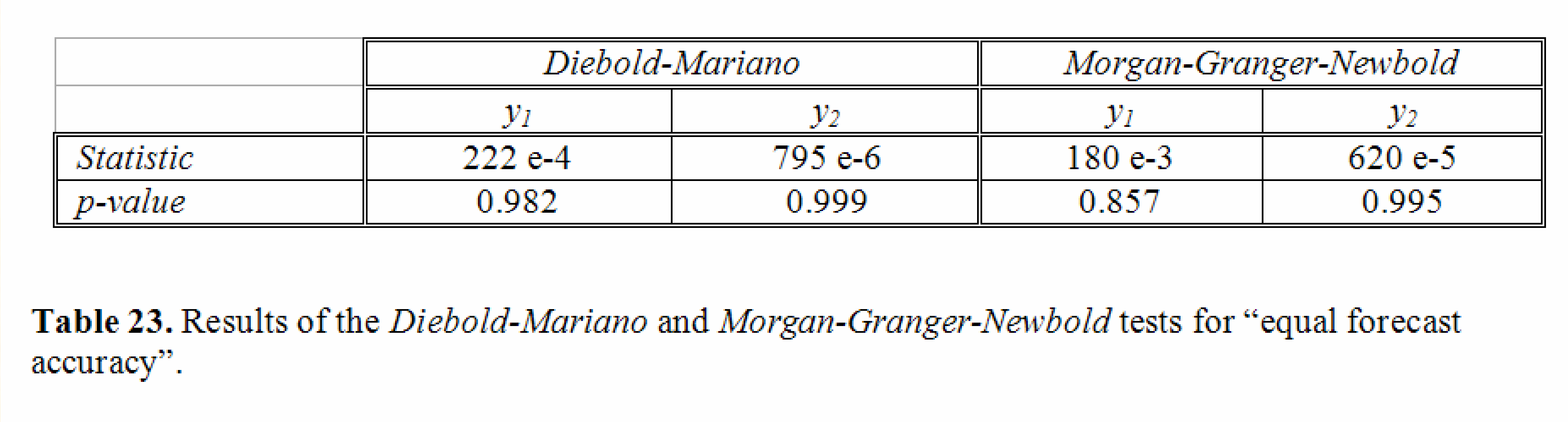
From 1000 observations

Recalculating the models (VEC, CCA)



	*MSPE (1.0 e-7 *)*		*MAPE(1.0 e-4 *)*	
	y_1	y_2	y_1	y_2
CCA	6589	3361	177	129
VAR	6565	3360	176	128
Random Walk	6591	3355	176	128

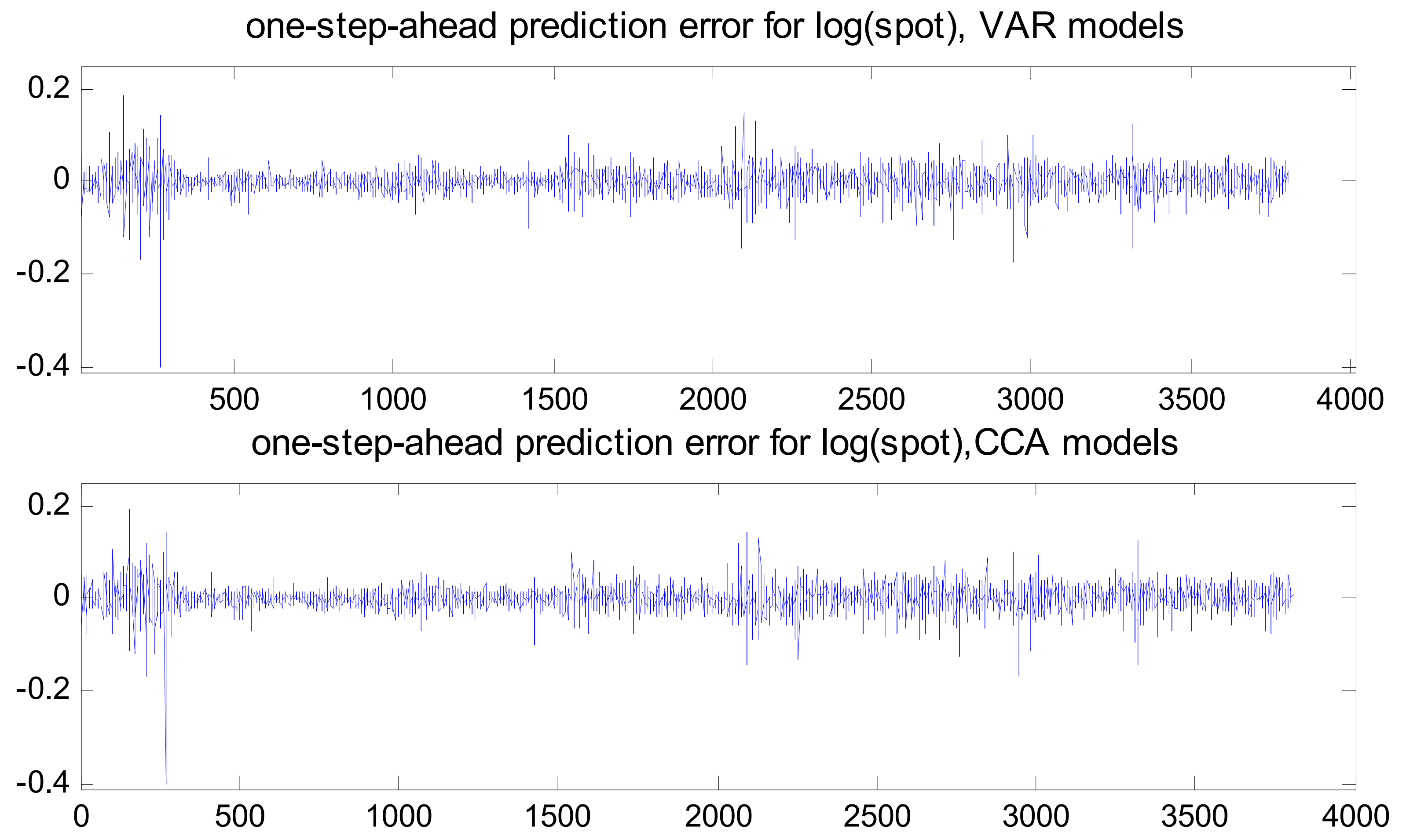
Table 22. Mean Square Prediction Error and Mean Absolute Prediction Error for the different models and series.



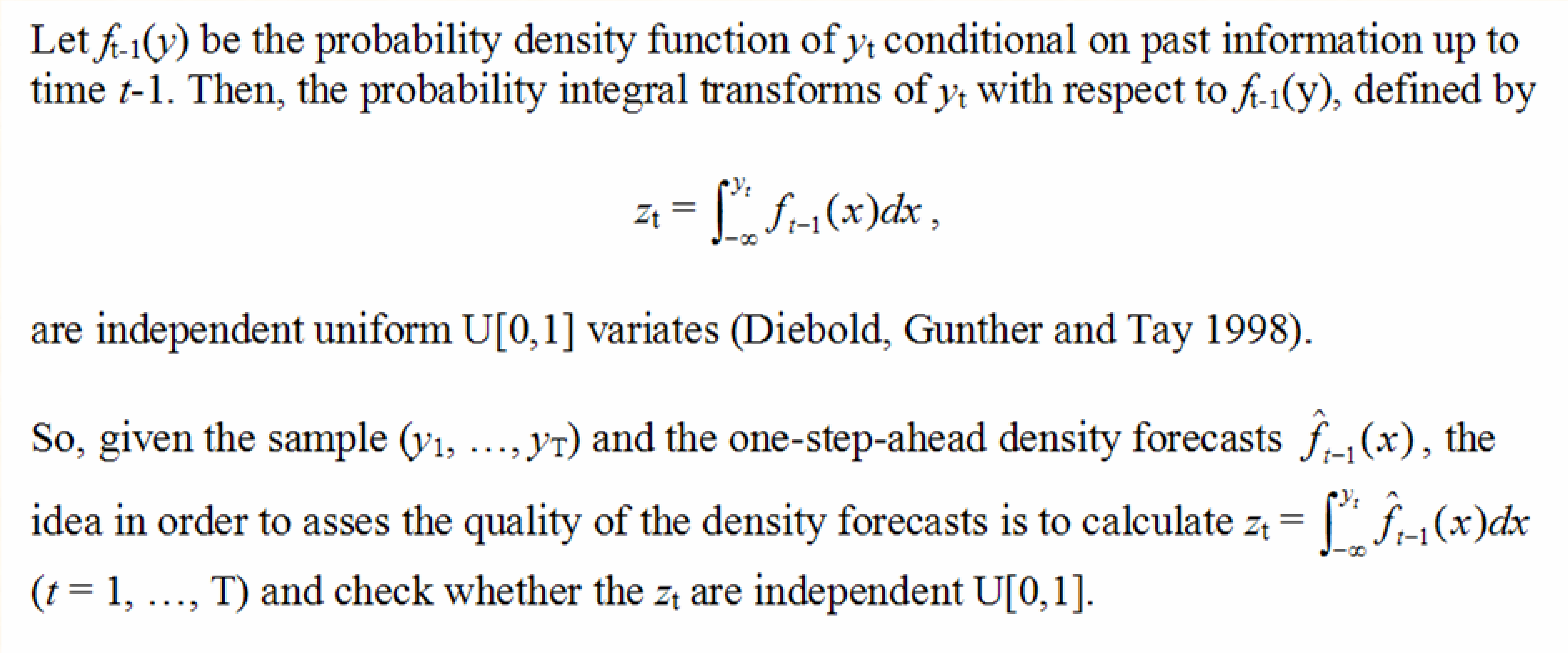
	Diebold-Mariano		*Morgan-Granger-Newbold*	
	y_1	y_2	y_1	y_2
Statistic	222 e-4	795 e-6	180 e-3	620 e-5
p-value	0.982	0.999	0.857	0.995

Table 23. Results of the *Diebold-Mariano* and *Morgan-Granger-Newbold* tests for "equal forecast accuracy".

4. Practical case



4. Practical case



Let $f_{t\text{-}1}(y)$ be the probability density function of y_t conditional on past information up to time t-1. Then, the probability integral transforms of y_t with respect to $f_{t\text{-}1}(y)$, defined by

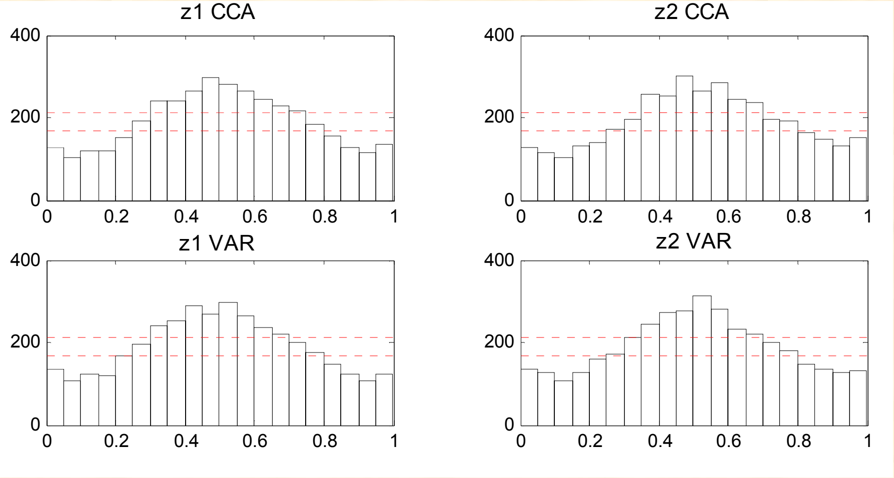
$$z_t = \int_{-\infty}^{y_t} f_{t-1}(x)dx\,,$$

are independent uniform U[0,1] variates (Diebold, Gunther and Tay 1998).

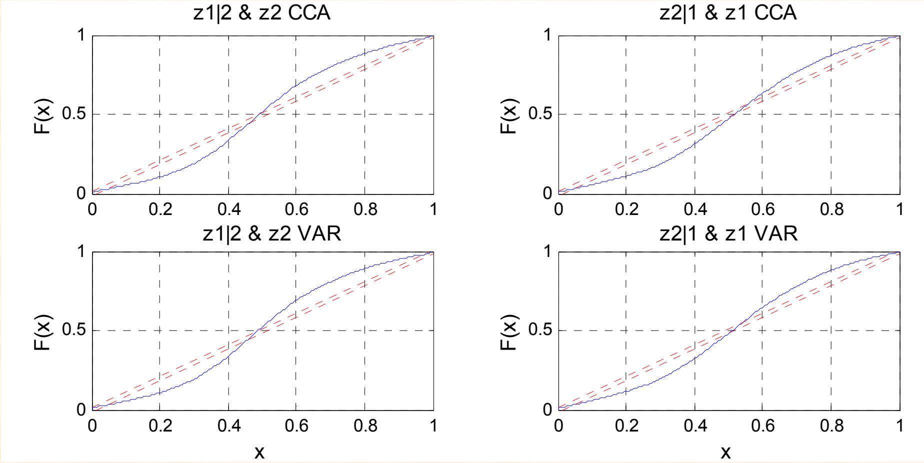
So, given the sample (y_1, …, y_T) and the one-step-ahead density forecasts $\hat{f}_{t-1}(x)$, the idea in order to asses the quality of the density forecasts is to calculate $z_t = \int_{-\infty}^{y_t} \hat{f}_{t-1}(x)dx$ (t = 1, …, T) and check whether the z_t are independent U[0,1].

4. Practical case

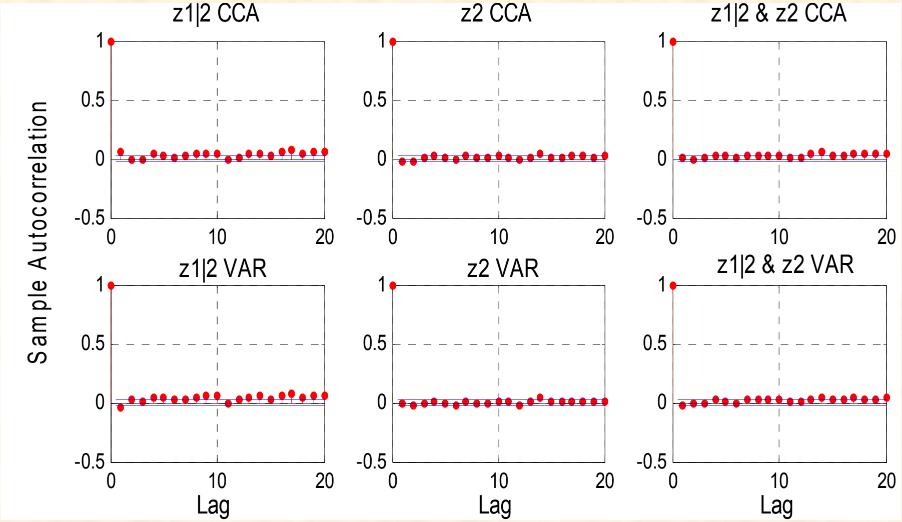
Bins, binomial test



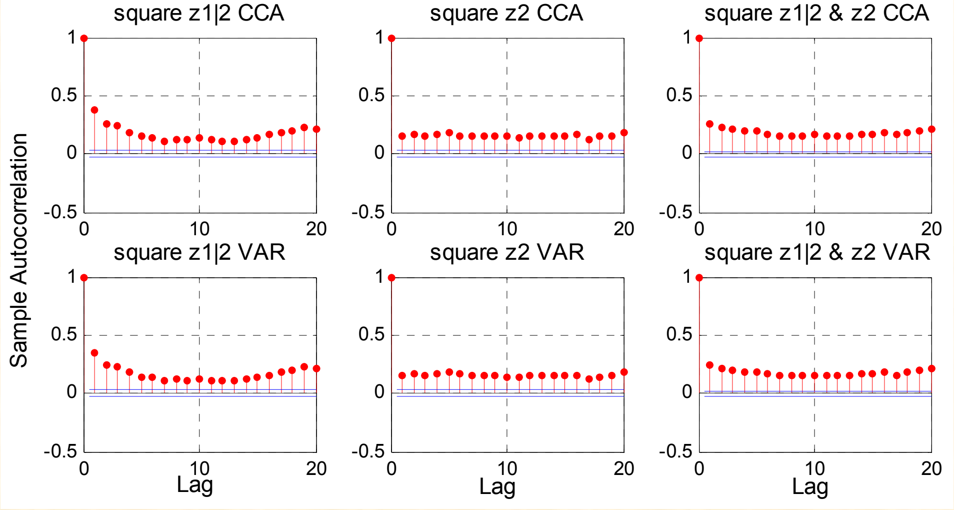
Empirical cdfs, K-S test



Sample autocorrelations of values



Sample autocorrelations of square values



4. Practical case

If we just take a window with the last two weeks to estimate the (changing) variance …