Analysis and forecast of the daily and intraday exchange rates with VAR, cointegration and using the Generalized Least Squares (GLS) + GLS analysis of simulated data.

Edy Zahnd
email: edy.zahnd@bluewin.ch

Journées suisses de la statistique, 27-29 août 2018,
Université de Zurich

Table of contents

PART 1: EXCHANGE RATES

Introduction

1. We study exchange rate pairs like EUR/USD and we use vector autoregressive models with cointegration relationship(s).

2. We use the close, high, low of prices for daily data.

3. In the VAR framework, we use several approaches: endogenous variables, general dummy variables (exogenous variables for intervention analysis), dummy variables for the day of week, interactions between these dummy variables and all indogenous variables, to account for the day-of-week effect, and the seasonal pattern.

4. We compare models with or without interactions,
5. We compare models for one or two pairs of currencies,
6. We compare the estimation over a long or a short period .
7. Then we compare 2 GLS methods

When we use a VAR/VEC model for the daily close, high and low prices of the EUR/USD, we get 3 equations including one or 2 cointegration relationships.

$$x_{1,t} = \Pi_1 \boldsymbol{\alpha}' \mathbf{x}_{\mathbf{t-1}} + a_{10} + \sum_{j=1}^{p} a_{11j} \Delta x_{1,t-j} + \sum_{j=1}^{p} a_{12j} \Delta x_{2,t-j} + \sum_{j=1}^{p} a_{13j} \Delta x_{3,t-j} + \sum_{j=0}^{r} b_{11j} z_{1,t-j} + \varepsilon_{1t} \qquad (1)$$

This is the first equation with only one cointegration relation: $\boldsymbol{\alpha}' \mathbf{x}_{\mathbf{t-1}}$, with its coefficient Π_1. In vector notation, we sum it up as follows:

$$x_{1,t} = \Pi_1 \alpha' x_{t-1} + a_{10} + a'_{1j} \Delta x_{1,t-j} + a'_{2j} \Delta x_{2,t-j} + a'_{3j} \Delta x_{3,t-j} + \sum_{j=0}^{r} b_{11j} z_{1,t-j} + \varepsilon_{1t} \qquad (2)$$

We can even write this in shorter form:

$$x_{1,t} = \Pi_1 \alpha' x_{t-1} + a_{10} + \sum_{i=1}^{3} a'_{ij}\, \Delta x_{t-j} + \sum_{j=0}^{r} b_{11j} z_{1,t-j} + \varepsilon_{1t}\,. \tag{3}$$

Models for the daily exchange rates

$$\begin{aligned} x^d{}_{1,t} &= \Pi_1 \alpha' x_{t-1} + a_{10} + \sum_{i=1}^{3} a'_{ij}\, \Delta x_{t-j} + \sum_{j=0}^{r} b_{11j} z_{1,t-j} + \varepsilon_{1t} + \\ &\sum_{i=1}^{4} D_{it} \left[\sum_{j=1}^{p} d_{11j} \Delta x_{1,t-j} + \sum_{j=1}^{p} d_{12j} \Delta x_{2,t-j} + \sum_{j=1}^{p} d_{13j} \Delta x_{3,t-j} \right] \end{aligned} \tag{4}$$

where one can see the dummies and their interactions with the variables, or more simply:

$$x^d{}_{1,t} = \Pi_1 \alpha' x_{t-1} + a_{10} + \sum_{i=1}^{3} a'_{ij}\, \Delta x_{t-j} + \sum_{j=0}^{r} b_{11j} z_{1,t-j} + \varepsilon_{1t} + \sum_{i=1}^{4} D_{it} \left[\sum_{j=1}^{p} d'_{1j} \Delta x_{t-j} \right] \tag{5}$$

where we used the vector d'_{1j} for the parameters d_{11j} , d_{12j} , d_{13j} , and the vector Δx_{t-j} for the variables x_{1t-j}, x_{2t-j}, x_{3t-j} in differences. Or even: $x^d{}_{1,t} = x_{1,t} + \sum_{i=1}^{4} D_{it} \left[\sum_{j=1}^{p} d'_{1j} \Delta x_{t-j} \right].$ (6)

This is just the first equation in the case with 3 variables. But we want to use 6 variables: the EUR/USD daily close, high and low, and the GBP/USD daily close, high and low.

Conclusion for VAR and cointegration
Results 1

TABLE 1
Model 1

TABLE 2
Model 2

Model 1 for EUR/USD alone with 2 COE with no interaction and no lag 5.			
Estimation over Jan 3, 2004 to Sept 8, 2010			
Currency	EUR/USD		
Price	CL	H	L
R2	0.0132	0.4796	0.4589
SER	87.602	59.374	56.314
RMSE of for.	**102.93**	**83.19**	**41.27**

Model 2 for EUR/USD and GBP/USD with interactions and lag 5			
Estimation over Jan 3, 2004 to Sept 8, 2010			
Currency	EUR/USD		
Price	CL	H	L
R2	0.0328	0.5016	0.4802
SER	87.417	58.566	55.629
RMSE of for.	**100.69**	**80.04**	**44.45**

TABLE 3
Neural network

Neural network model: three 1-step-ahead forecasts over 3 days only: Sept 9, 10 and 13			
Estimation: a short period up to Sept. 8, 2010			
Currency	EUR/USD		
Price	CL	H	L
RMSE of for.	**-**	**95.07**	**40.44**

Results 2

1) The model with interactions and a lag 5 is better than without both. R-square and SER are smaller and the RMSE (root of mean squared error) is lower, meaning that the forecast is better.

2) This model outperforms a NN model for the value of the high, but the NN is better for the low.

3) When the estimation is performed on a shorter sample, the SER are smaller, the R-squared are higher, and the forecast is often better.

4) The models with cointegration lead to better results than simple VARs.

Daily exchange rates and GLS, 2 methods

1) We used a diagonal G matrix
2) in method one, we estimate the variances of the residuals of each day by the squared differences between the high H and the low L. If H-L=0, we use 0.005 instead.
3) in method 2, we estimate the variances by the squared residuals provided by the OLS estimation.

Conclusion for GLS

1) As theory says, the variance of GLS estimators is smaller.
2) Our method 1 worked and showed some improvements in the forecasts compared with OLS, but
3) The second method works better
4) The estimates of the parameters in both methods may be slightly biased, but the variance drops in a substantial way, so that the forecast improves.

Improved parameter estimation and predictions using the Generalized Least Squares in the presence of heteroscedasticity

Edy Zahnd
email: edy.zahnd@bluewin.ch

Journées suisses de la statistique,
Université de Zurich, 27.8.2018

Least Squares and Generalized Least Squares

→ consider a linear regression model $Y = X\beta + \varepsilon$ with $\mathrm{Var}(\varepsilon) = \sigma^2 G$ where :

- Y is a vector of n observations of a response variable
- X is a design matrix of dimension $n \times (p+1)$
  (containing a column of ones and the data on p predictors)
- β is a vector of $p+1$ unknown parameters (one intercept and p slopes)
- ε is a vector of n residuals
- G is a $n \times n$ variance-covariance matrix (and σ^2 is a scalar)

→ the Least Squares (LS) estimate of β is obtained as :

$$\widehat{\beta}_{LS} = (X'X)^{-1}X'Y$$

→ the Generalized Least Squares (GLS) estimate of β is obtained as :

$$\widehat{\beta}_{GLS} = (X'G^{-1}X)^{-1}X'G^{-1}Y$$

Variance of LS and GLS estimates

→ both the LS and GLS estimates are unbiased with (e.g. Kennedy, 1998) :

$$\mathrm{Var}(\widehat{\beta}_{LS}) = \sigma^2 \cdot (X'X)^{-1} X' G X (X'X)^{-1}$$

and :

$$\mathrm{Var}(\widehat{\beta}_{GLS}) = \sigma^2 \cdot (X' G^{-1} X)^{-1}$$

→ with homoscedastic and independent residuals, one has $G = I$ (the identity matrix) and :

$$\mathrm{Var}(\widehat{\beta}_{LS}) = \mathrm{Var}(\widehat{\beta}_{GLS}) = \sigma^2 \cdot (X'X)^{-1}$$

→ with heteroscedastic and/or non-independent (correlated) residuals, the diagonal elements of $\mathrm{Var}(\widehat{\beta}_{GLS})$ are smaller than the diagonal elements of $\mathrm{Var}(\widehat{\beta}_{LS})$

→ GLS more efficient than LS

→ Problem : one does not know G in practice !

Two-stage GLS estimate

→ G is a diagonal matrix if one assumes independent residuals

→ the diagonal elements of $\sigma^2 G$ can be estimated non-parametrically by $\widehat{\varepsilon}_t^2$, the estimated residuals obtained from $\widehat{\beta}_{LS}$ $(t = 1, \cdots, n)$, yielding :

$$\widehat{G} = \begin{pmatrix} \widehat{\varepsilon}_1^2 & 0 & \cdots & 0 \\ 0 & \widehat{\varepsilon}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widehat{\varepsilon}_n^2 \end{pmatrix}$$

→ one may then consider the two-stage GLS estimate defined as :

$$\widehat{\beta}_{TSGLS} = (X'\widehat{G}^{-1}X)^{-1}X'\widehat{G}^{-1}Y$$

→ this is actually a special case of "Weighted Least Squares", where each observation t is weighted (in a second stage) by $1/\widehat{\varepsilon}_t^2$ (obtained in a first stage)

→ a similar procedure has been proposed by White (1980) to get consistent estimates for the standard errors of the elements of $\widehat{\beta}_{LS}$ in the case of heterosckedastic errors (also used for a test of heteroscedasticity)

Excerpt of Leamer (1988) (Three things that bother me)

“I see a great increase in the use of asymptotic theory to justify our statistical analyses. This worries me. ‘White-washing’ heteroscedasticity is an example. By ‘White-washing’ I mean correcting the standard errors, but leaving the regression coefficient estimates unchanged, a procedure that has been endorsed by White (1980). I know that this surprising method has an appealing asymptotic justification, but this treatment for the pathology of heteroscedasticity is very different from the traditional medicine. Traditionally, one was expected to think about the form that heteroscedasticity might take, and to use an appropriately chosen set of weights to form weighted least squares that corrected not just the standard errors but the regression coefficients as well. Surely, for some data sets it is a mistake not to adjust the estimates and to adjust only the standard errors. One can get away with adjusting only the standard errors asymptotically, but there is no theorem of which I am aware that justifies this is a finite data set.”

→ in this presentation, we are interested to quantify the possible gain achieved when using TSGLS instead of LS on the precision of parameter estimation and on the quality of prediction in the presence of heteroscedasticity in finite data sets

Quantifying the gain on the precision of parameter estimation

→ measure of % gain on precision of parameter estimation of the jth element β_j of β :

$$\text{GAIN EST} = 100 \cdot \frac{\text{MSE}(\widehat{\beta}_{LS,j}) - \text{MSE}(\widehat{\beta}_{TSGLS,j})}{\text{MSE}(\widehat{\beta}_{LS,j})}$$

where $\widehat{\beta}_{LS,j}$ and $\widehat{\beta}_{TSGLS,j}$ are the jth elements of $\widehat{\beta}_{LS}$ and $\widehat{\beta}_{TSGLS}$, and where the Mean Square Error (MSE) of an estimate $\widehat{\beta}_j$ of β_j is defined as

$$\text{MSE}(\widehat{\beta}_j) = \text{E}[(\widehat{\beta}_j - \beta_j)^2] = \text{Bias}^2(\widehat{\beta}_j) + \text{Var}(\widehat{\beta}_j)$$

(equals to $\text{Var}(\widehat{\beta}_j)$ for unbiased estimates)

→ interpretable as % of improvement when using TSGLS instead of LS

→ positive values favor TSGLS, negative values favor LS

→ can be estimated via simulations with various G

Quantifying the gain on the quality of prediction

$\rightarrow$ measure of % gain on quality of prediction for *m* future observations :

$$\text{GAIN PRED} = 100 \cdot \frac{\text{MSPE}_{LS} - \text{MSPE}_{TSGLS}}{\text{MSPE}_{LS}}$$

where the Mean Square Prediction Error (MSPE) achieved using an estimate $\widehat{\beta}$ of β to predict m future observations is defined as :

$$\text{MSPE}(\widehat{\beta}) = \text{E}\left[\frac{1}{m}\sum_{k=1}^{m}\widehat{\varepsilon}_k^2\right]$$

(where $\widehat{\varepsilon}_k$ is the difference between the actual value and the predicted value using $\widehat{\beta}$ for the future observation Y_k, $k = 1, \cdots, m$)

$\rightarrow$ interpretable as % of improvement when using TSGLS instead of LS

$\rightarrow$ positive values favor TSGLS, negative values favor LS

$\rightarrow$ can be estimated via simulations with various G

Simulations design

→ linear model with a single predictor (time) : $Y_t = \alpha + \beta \cdot t + \varepsilon_t$ $(t = 1, \cdots, n+m)$ with $\varepsilon_t = \rho \cdot \varepsilon_{t-1} + \sqrt{1-\rho^2} \cdot u_t$, where the u_t are i.i.d normal "innovations" with mean 0 and variance σ_t^2 defined below (without loss of generality, we set $\alpha = 0$)

→ we considered various combinations of the following factors :

- sample size $n = 50, 100, 1000$
- number of future observations $m = 1, 10$
- % of explained variance by the trend $R^2 = 0.3, 0.8$ where :

$$R^2 = \frac{\beta^2 \text{Var}(t)}{\beta^2 \text{Var}(t) + E(\sigma_t^2)}$$

where $\sigma_t^2 = \text{Var}(\varepsilon_t)$, which is achieved by taking :

$$\beta = \sqrt{\frac{R^2}{1-R^2} \cdot \frac{E(\sigma_t^2)}{\text{Var}(t)}}$$

Note : we have $\text{Var}(t) \approx (n+m)^2/12$ whereas we took $E(\sigma_t^2) = 5$

Note : this factor R^2 had actually no influence on our results

Simulations design

- heteroscedasticity : diagonal element σ_t^2 of $\sigma^2 G$ taken as $(t = 1, \cdots, n+m)$:

(H1) $\sigma_t^2 = 5$ (homoscedasticity)

(H2) $\sigma_t^2 =$ a random number between 0 and 10

(H3) $\sigma_t^2 = (10t)/(n+m+1)$ (linear trend)

(H4) $\sigma_t^2 = 5 + 5\sin((4\pi t)/(n+m))$ (sinusoidal trend)

(H5) $\sigma_t^2 =$ either 1 or 9 (with probability 50%-50%)

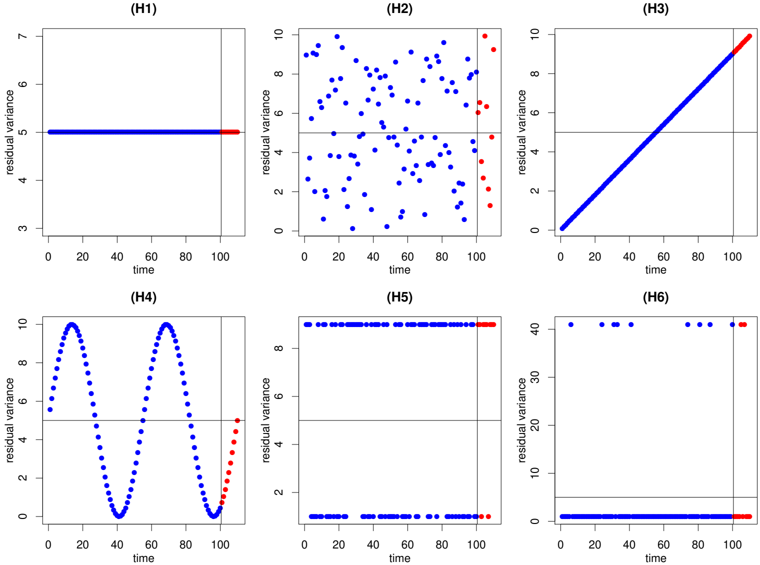
(H6) $\sigma_t^2 =$ either 1 (probability 90%) or 41 (probability 10%)
(mimics the presence of 10% of outliers)

Note : the average (expectation) of σ_t^2 over $t = 1, \cdots, n+m$ is 5 in each case

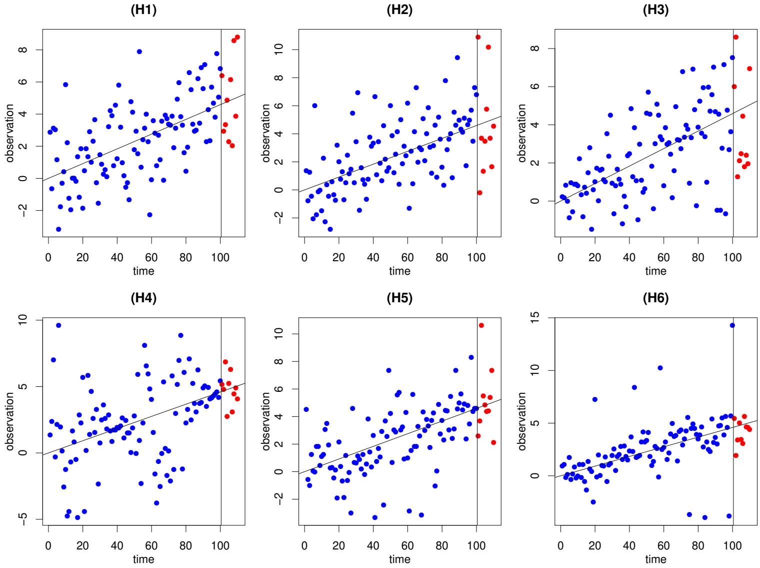
- autocorrelation of residuals : $\rho = \mathrm{Cor}(\varepsilon_t, \varepsilon_{t-1}) = 0, 0.5$ (two different models)

$\rightarrow$ GAIN EST (with respect to the estimation of the slope β) and GAIN PRED were estimated based on 5000 simulations (in each considered setting)

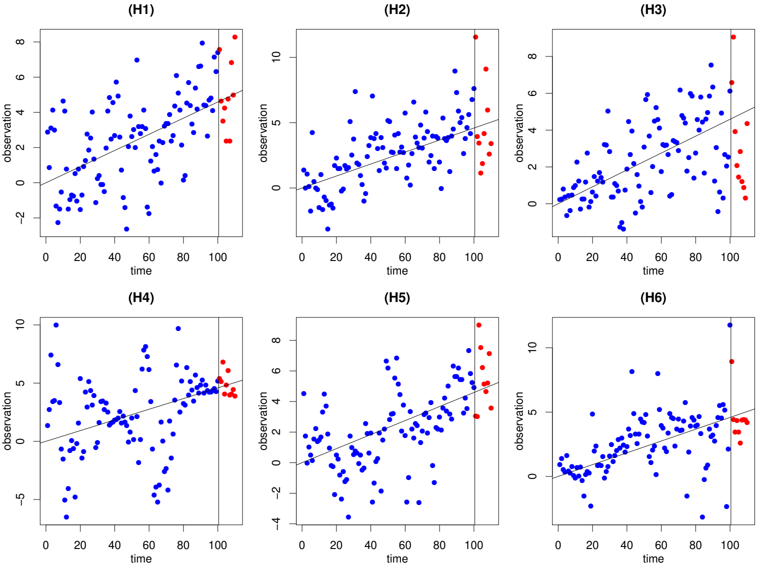
The six residual variance functions considered



Example of simulated data ($n = 100,\ m = 10,\ R^2 = 0.3,\ \rho = 0$)



Simulated data with autocorrelation ($n = 100$, $m = 10$, $R^2 = 0.3$, $\rho = 0.5$)



Results $(n = 50)$

σ_t^2	ρ	$m = 1$ GAIN EST	$m = 1$ GAIN PRED	$m = 10$ GAIN EST	$m = 10$ GAIN PRED
(H1)	0	-2.9	-0.3	-3.4	-0.4
(H2)	0	2.4	0.3	3.1	0.3
(H3)	0	1.3	-0.1	2.0	0.0
(H4)	0	8.7	0.5	1.6	0.5
(H5)	0	4.3	0.3	4.5	0.5
(H6)	0	7.6	0.8	7.3	0.6
(H1)	0.5	-1.4	-0.5	-0.8	-0.2
(H2)	0.5	0.3	-0.1	0.4	0.1
(H3)	0.5	2.6	-0.2	3.5	0.1
(H4)	0.5	6.6	1.3	1.5	1.1
(H5)	0.5	2.0	0.2	1.8	0.4
(H6)	0.5	7.0	0.7	6.5	1.3

Results ($n = 100$)

σ_t^2	ρ	$m = 1$ GAIN EST	$m = 1$ GAIN PRED	$m = 10$ GAIN EST	$m = 10$ GAIN PRED
(H1)	0	−2.1	−0.2	−2.5	−0.1
(H2)	0	1.9	0.1	2.1	0.1
(H3)	0	1.5	0.0	1.2	0.0
(H4)	0	6.5	0.3	8.0	0.7
(H5)	0	1.9	0.1	2.4	0.1
(H6)	0	3.5	0.1	3.6	0.1
(H1)	0.5	−0.5	−0.2	−1.0	−0.1
(H2)	0.5	0.4	0.1	0.4	0.1
(H3)	0.5	1.7	−0.1	2.3	0.1
(H4)	0.5	4.6	0.3	5.4	1.3
(H5)	0.5	1.1	0.0	1.0	0.1
(H6)	0.5	3.2	0.2	3.1	0.3

Results ($n = 1000$)

σ_t^2	ρ	$m = 1$ GAIN EST	$m = 1$ GAIN PRED	$m = 10$ GAIN EST	$m = 10$ GAIN PRED
(H1)	0	0.0	0.0	−0.3	0.0
(H2)	0	0.3	0.0	0.3	0.0
(H3)	0	0.3	0.0	0.4	0.0
(H4)	0	1.5	0.0	1.6	0.0
(H5)	0	0.1	0.0	0.1	0.0
(H6)	0	0.2	0.0	0.3	0.0
(H1)	0.5	−0.2	0.0	−0.1	0.0
(H2)	0.5	0.0	0.0	0.1	0.0
(H3)	0.5	0.4	0.0	0.4	0.0
(H4)	0.5	1.1	0.0	1.1	0.0
(H5)	0.5	0.2	0.0	0.0	0.0
(H6)	0.5	0.3	0.0	0.3	0.0

Simulations with an AR(1) process

→ model : $Y_t = \alpha + \beta \cdot Y_{t-1} + \varepsilon_t$ $(t = 1, \cdots, n+m)$ where the ε_t are i.i.d normal with mean 0 and variance σ_t^2 as defined above (we set $\alpha = 0$ and $\beta = 0.5$)

→ $n = 50$

	$m = 1$		$m = 10$	
σ_t^2	GAIN EST	GAIN PRED	GAIN EST	GAIN PRED
(H1)	-3.9	-0.3	-4.4	-0.1
(H2)	2.0	0.0	1.5	0.1
(H3)	-1.1	-0.1	-1.4	0.0
(H4)	0.7	0.2	-1.1	0.5
(H5)	4.8	0.3	4.6	0.1
(H6)	5.8	0.5	5.6	0.2

Simulations with an AR(1) process

$\rightarrow$ $n = 100$

σ_t^2	$m = 1$ GAIN EST	$m = 1$ GAIN PRED	$m = 10$ GAIN EST	$m = 10$ GAIN PRED
(H1)	-1.7	-0.1	-2.2	-0.1
(H2)	1.8	0.0	1.4	0.1
(H3)	-1.2	0.1	-0.6	0.0
(H4)	0.6	0.1	0.2	0.2
(H5)	2.7	0.1	1.6	0.0
(H6)	4.1	0.1	4.2	0.1

$\rightarrow$ $n = 1000$

σ_t^2	$m = 1$ GAIN EST	$m = 1$ GAIN PRED	$m = 10$ GAIN EST	$m = 10$ GAIN PRED
(H1)	-0.2	0.0	-0.3	0.0
(H2)	0.6	0.0	0.5	0.0
(H3)	-0.2	0.0	0.2	0.0
(H4)	0.1	0.0	-0.1	0.0
(H5)	0.2	0.0	0.3	0.0
(H6)	0.7	0.0	0.5	0.0

Some conclusions

→ for homoscedastic errors, TSGLS logically (slightly) worse than LS

→ for heteroscedastic errors, TSGLS often better than LS, the improvement being more important for estimation than for prediction

→ in particular, TSGLS consistently better than LS in a model with outliers (H6)

→ difference between TGLS and LS more important for small n than for large n

→ for $n = 1000$, there is virtually no difference between TSGLS and LS for prediction

→ Note : the gain with respect to estimation is not always useful in practice (since the standard errors of TSGLS estimates may be be difficult to estimate)

→ a gain with respect to prediction would be useful in practice, but the results are sometimes a bit disappointing (gain often smaller than 1%)

Some references

- Chatfield C (1989). *Analysis of Time Series : An Introduction*. Chapman and Hall.
- Fox J (2008). *Applied Regression Analysis and Generalized Linear Models*. Sage.
- Fox J, Weisberg S (2011). *Time Series Regression on Generalized Least Squares in R*. Appendix to *An R Companion to Applied Regression*, second edition, Sage.
- Greene WH (2003). *Econometric Analysis*, fifth edition, Prentice Hall.
- Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee TC. *The Theory and Practice of Econometrics*, second edition, Wiley.
- Kennedy P (1998). *A Guide to Econometrics*, fourth edition, Blackwell.
- Leamer E (1988). Three things that bother me. *Economic Record*, 64, 331-335.
- White H (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838