USING AUTOMETRICS

Genaro Sucarrat

(Department of Economics, UC3M)

http://www.eco.uc3m.es/sucarrat/

March 25, 2008

- The basics of OxMetrics 5:

  → Loading, editing and transforming data, creating "special" series (cointegration relations, trends, etc.)

- An overview of Autometrics:

  → Key concepts and characteristics

- Single-equation modelling with Autometrics

  → Formulation, Advanced Autometrics settings, fixing variables, example (2007 Econometric Game, Q1)

- Multiple-equation modelling with Autometrics

  → Formulation, fixing variables, example (2007 Econometric Game, Q2)

OxMetrics basics:

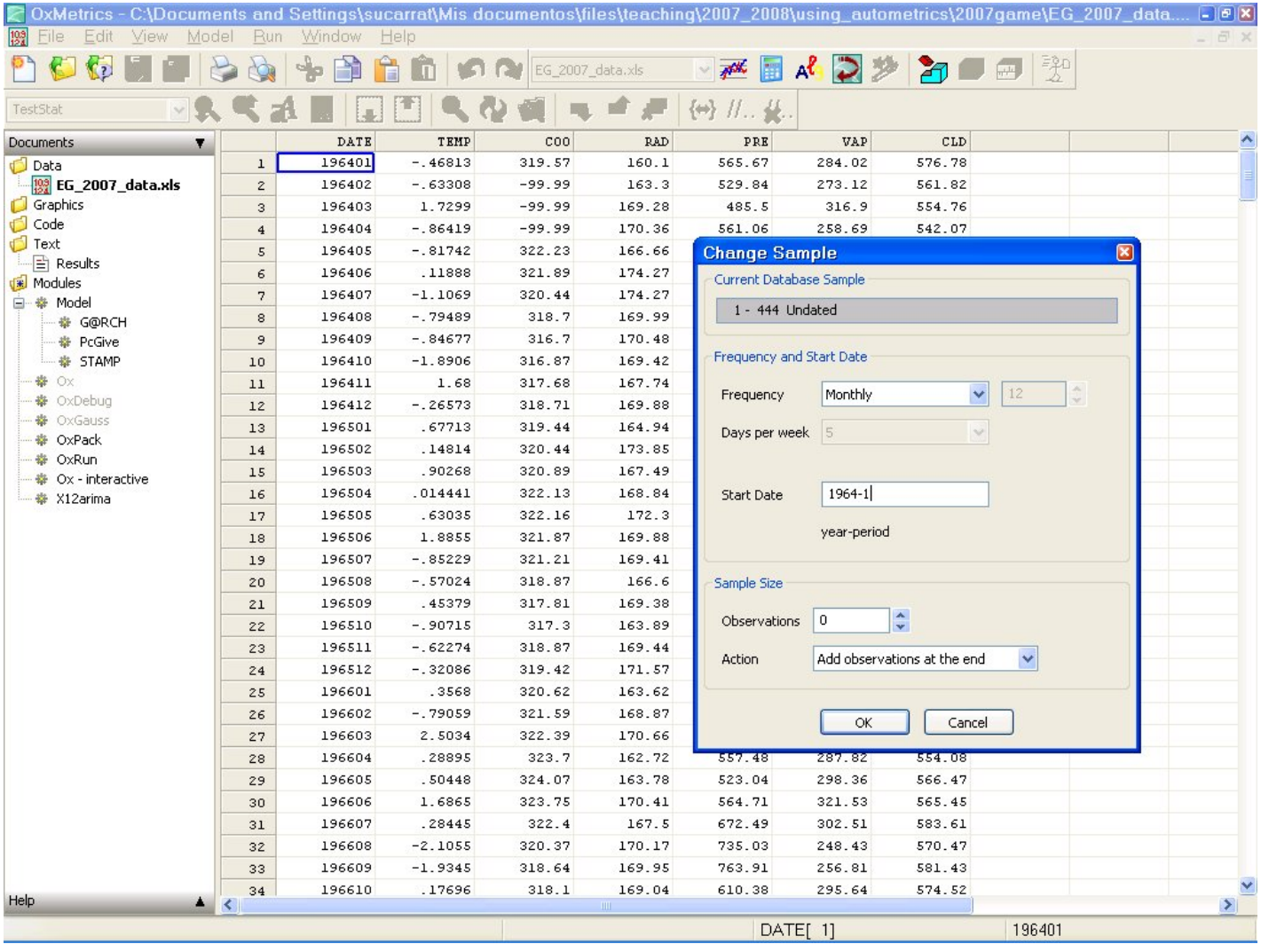
- Load data: File → Open, etc.
- Edit sample/dates: Edit → Change Sample
- Missing values (recommendation): Set to "missing" by double-clicking the data cell in question
- Graph series: Model → Graphics (or click on the graphics button) → Actual series or All plot types
- Transform data (algebra feature):

  Edit → Algebra (or click on the "Alg" button) → "Code", e.g. "DCOO = diff(COO,1);" → Run (→ Done)

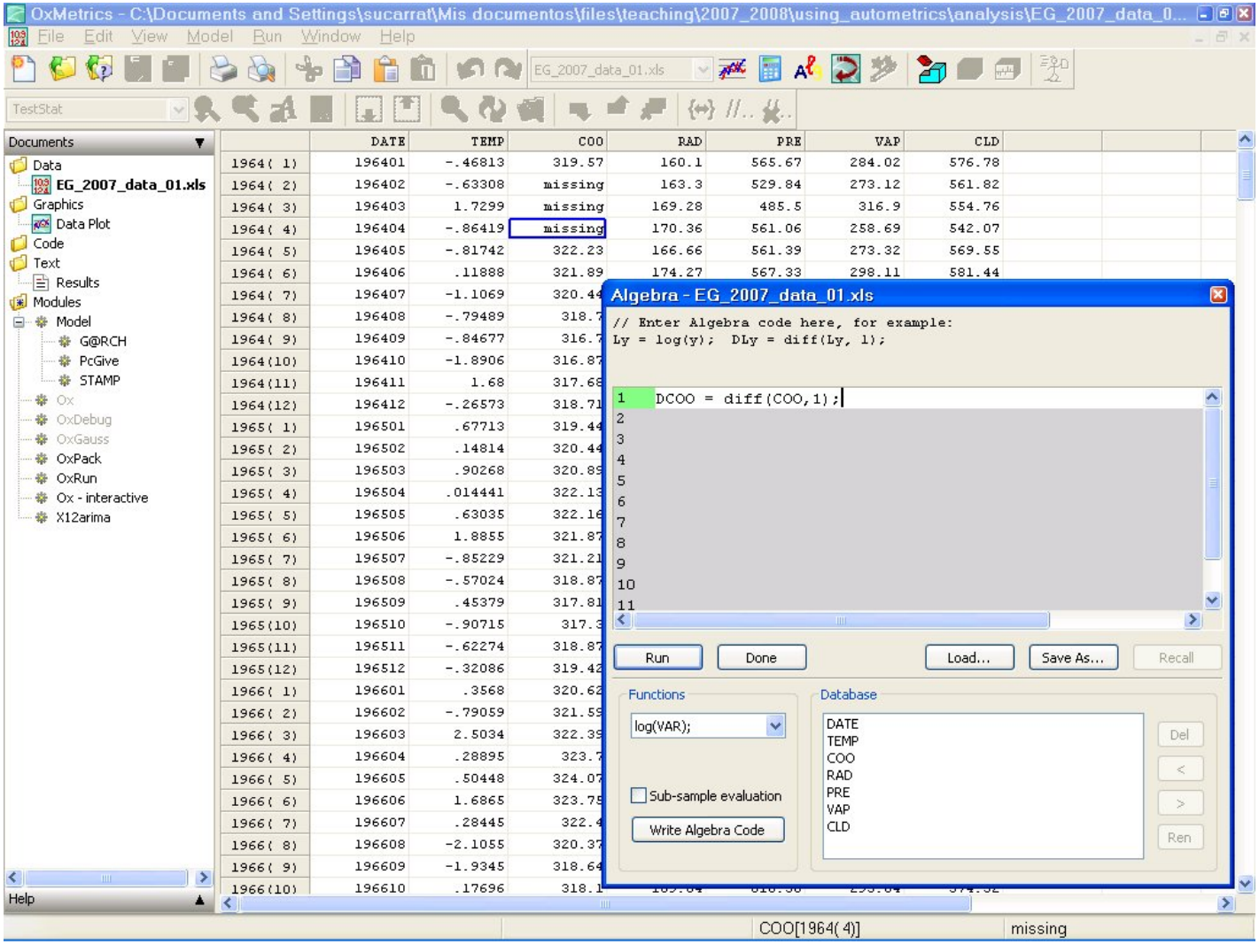
  (NOTE: Case sensitivity in variable names!)

- Create special series (calculator feature): Model → Calculator (or click on the calculator button) → ...

- *Example*. Edit dates (2007 Econometric Game Case):



- *Example*. Create a differenced series:



An overview of Autometrics:

• The objective of Autometrics is to automate *General-to-Specific* (Gets) *multiple path* simplification search of a *data coherent*, *General Unrestricted Model* (GUM) in the form of a linear OLS/IV estimable regression (or regressions)

• Default definition of data-coherency: Stable parameters and Gaussian, serially uncorrelated, homoscedastic errors. NOTE: These assumptions can be relaxed through the "Advanced Autometrics settings", and if the GUM fails one or several diagnostic checks Autometrics proceeds anyway

• GUM: A general model (advice: Not too general!) that includes the variables and lags that are believed to possibly have an impact

• Further reading: Doornik and Hendry (2007a, pp. 70-77), Hendry and Krolzig (2001) (Autometrics is an evolution of PcGets)

- Key benefits of Gets modelling:

  $\longrightarrow$ Estimation and inference while controlling for the impact from other variables: Spurious variables are more likely to be excluded, parameter estimates are "more correct"

  $\longrightarrow$ Gets modelling leads to a parsimonious, explanatory model particularly useful for scenario analysis (policy analysis, counterfactual analysis and conditional forecasting)

- Originally, the main disadvantage of Gets modelling was:

  $\longrightarrow$ Resource demanding and time consuming if properly implemented

- Solution: Automated Gets

  $\longrightarrow$ Hoover and Perez (1999), PcGets (Hendry and Krolzig 2001), Autometrics (Doornik 2007, Doornik and Hendry 2007a)

What does multiple path Gets specification search consist in?

• Sequential deletion of insignificant variables (significance level can be user-specified), while checking data-coherency for each variable deletion (variable-deletion that induces data-incoherency, say, serially correlated errors, is not undertaken)

• Multiple path Gets search typically leads to several terminal models; Autometrics either applies Gets on their union or chooses among them by means of information criteria (can be user-specified; keyword: "Tie-breaker")

Single-equation modelling with Autometrics. *Example*: 2007 Econometric Game, Question 1

- My GUM:

$$\Delta COO_t = b_0 + b_1 \Delta COO_{t-1} + b_2 \Delta COO_{t-2} + \sum_{j=1}^{11} c_j d_{j,t} + e_t$$

- The specific model proposed by Autometrics using the default options:

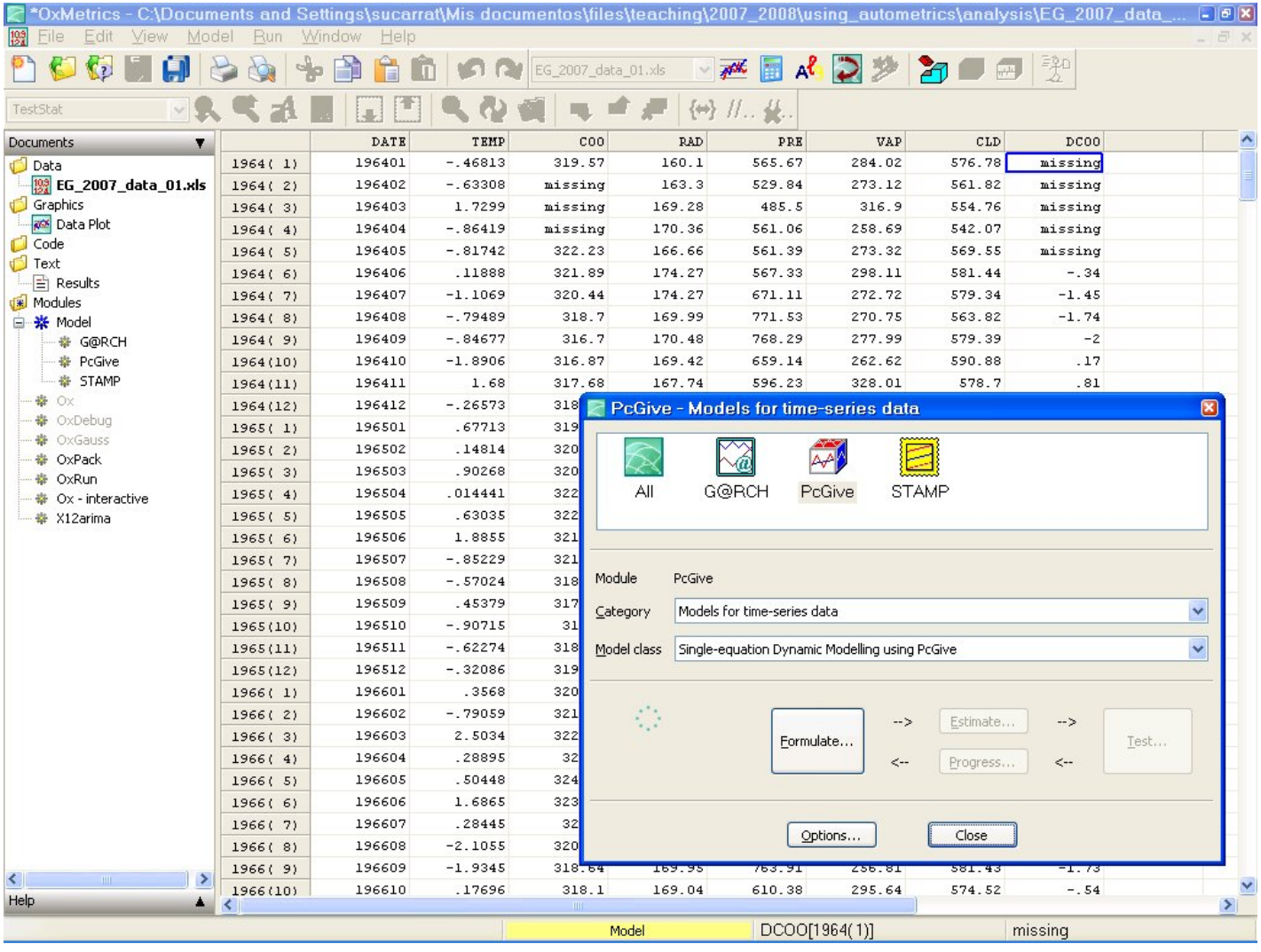
$$\Delta COO_t = b_0 + b_1 \Delta COO_{t-1} + \sum_{j=1}^{3} c_j d_{j,t} + \sum_{j=5}^{11} c_j d_{j,t} + e_t$$

- Unfortunately, the missing values cannot be estimated by means of the level representation

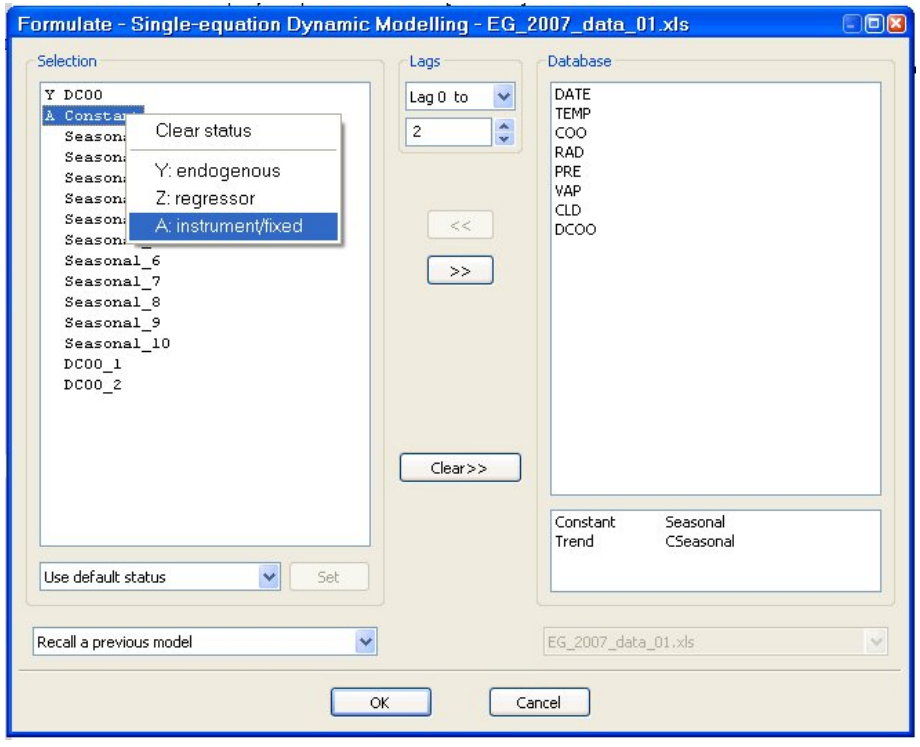
$$COO_t = b_0 + (1 + b_1) COO_{t-1} + b_1 COO_{t-2} + \sum_{j=1}^{3} c_j d_{j,t} + \sum_{j=5}^{11} c_j d_{j,t} + e_t$$

so a different specification is needed with, say, no lags of ΔCOO_t

- Formulate a model: (Model →) PcGive → Category: "Models for time series data" → Model class: "Single-equation dynamic modelling using PcGive" → "Formulate"



- Specify model:



USEFUL FEATURE: Fixing regressors (that is, preventing Autometrics from deleting them). Select the regressors to fix → Right-click mouse → A: instrument/fixed. NOTE: This defines instrument if IV is used instead of OLS

- Selected Autometrics options:

  → Target size: Significance level

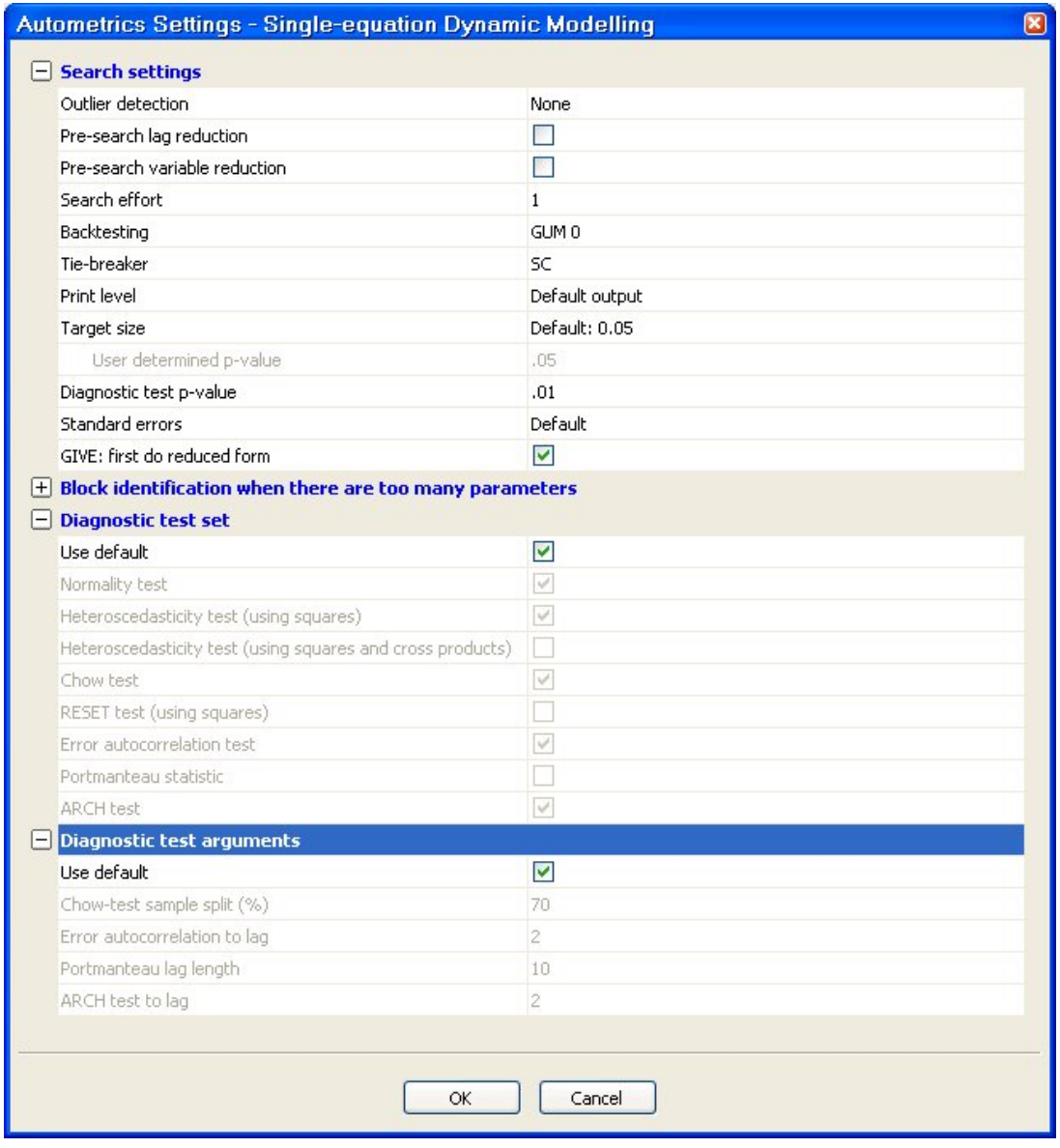
  → Outlier detection: Neutralises large residuals in the GUM by means of impulse dummies

  → Pre-search lag reduction: Speeds up simplification; GENERAL ADVICE: Turn off!

  → Advanced Autometrics settings: In general, use only if default settings and options are unsatisfactory

  → Recursive graphics: TURN ON! Slows down computation (slightly), but enables some very useful stability diagnostics

- Advanced Autometrics settings:



- Specific model proposed by Autometrics:

```
EQ( 2) Modelling DCOO by OLS
       The dataset is: C:\Documents and Settings\sucarrat\Mis documentos\files\teaching
       The estimation sample is: 1964(8) - 2000(12)

                Coefficient  Std.Error  t-value  t-prob Part.R^2
DCOO_1             -0.213999    0.04512    -4.74  0.0000   0.0503
Seasonal           -0.199927    0.05880    -3.40  0.0007   0.0265
Seasonal_1         -0.507123    0.05851    -8.67  0.0000   0.1502
Seasonal_2         -0.451573    0.06034    -7.48  0.0000   0.1165
Seasonal_4         -0.707806    0.05885    -12.0  0.0000   0.2540
Seasonal_5          -1.94870    0.06413    -30.4  0.0000   0.6848
Seasonal_6          -3.02784    0.09527    -31.8  0.0000   0.7038
Seasonal_7          -3.77271     0.1275    -29.6  0.0000   0.6733
Seasonal_8          -3.69728     0.1505    -24.6  0.0000   0.5868
Seasonal_9          -1.89487     0.1424    -13.3  0.0000   0.2941
Seasonal_10         -0.176490    0.07692    -2.29  0.0222   0.0122
Constant      U      1.50663    0.06042     24.9  0.0000   0.5940

sigma               0.286776  RSS                34.9520821
R^2                 0.946555  F(11,425) =   684.3 [0.000]**
log-likelihood      -68.1549  DW                      2.05
no. of observations      437  no. of parameters         12
mean(DCOO)          0.112334  var(DCOO)            1.49654

AR 1-7 test:      F(7,418)  =   1.0154 [0.4196]
ARCH 1-7 test:    F(7,411)  =  0.78778 [0.5979]
Normality test:   Chi^2(2)  =   2.5460 [0.2800]
Hetero test:      F(12,412) =  0.85756 [0.5908]
Hetero-X test:    F(22,402) =   1.2915 [0.1715]
RESET test:       F(1,424)  =0.00067988 [0.9792]
```

Further diagnostic tests:

- Residuals graphs: Model $\rightarrow$ Test $\rightarrow$ Graphical analysis $\rightarrow \ldots$
- User specified residuals tests: Model $\rightarrow$ Test $\rightarrow$ Test $\rightarrow \ldots$
- Recursive graphics (VERY useful!): Model $\rightarrow$ Test $\rightarrow$ Recursive graphics $\rightarrow \ldots$

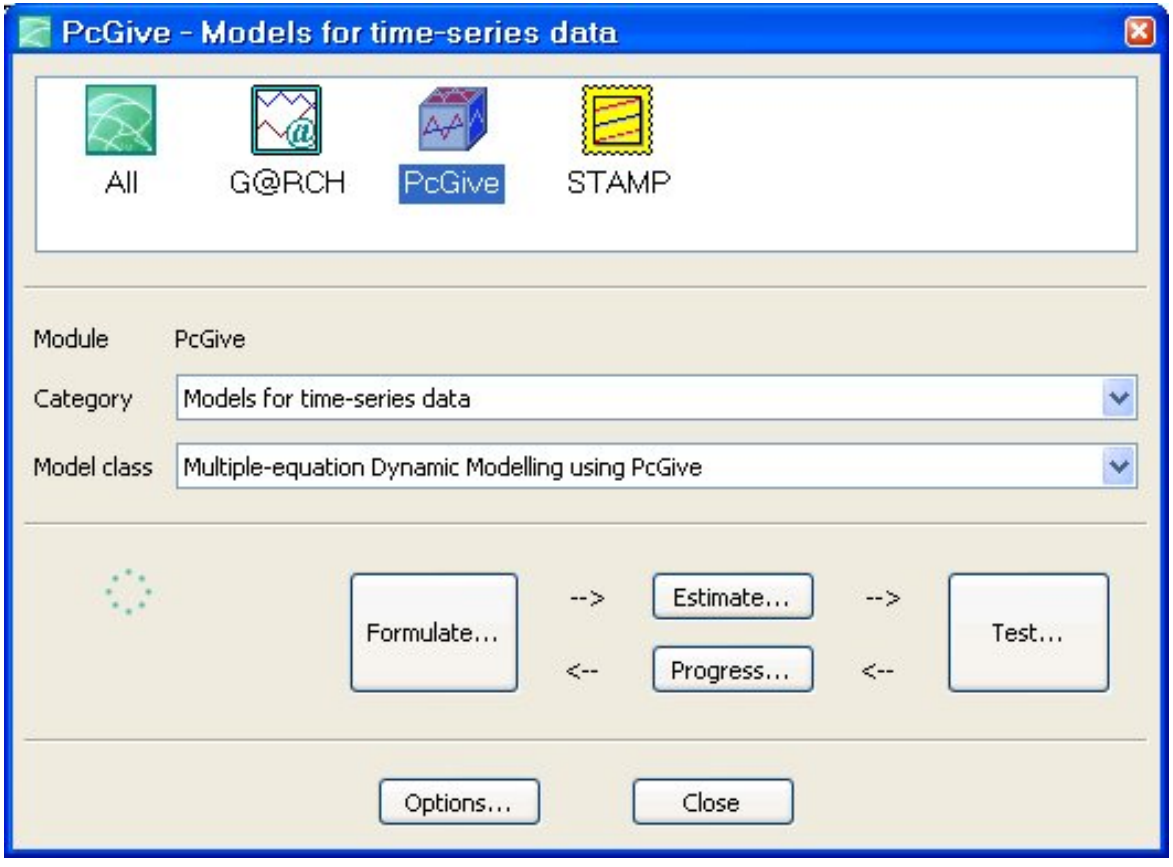
Multiple-equation modelling with Autometrics: Two possibilities

• Seemingly Unrelated Regression (SUR) using OLS/IV, that is, single-equation modelling of each equation separately (requires stationarity of regressors)

• Simultaneous variable deletion (or non-deletion) across equations using vector diagnostic tests but estimation still by OLS (does not require stationarity of regressors), see Doornik and Hendry (2007b, pp. 29-31). (NOTE: IV estimation not available with this strategy)

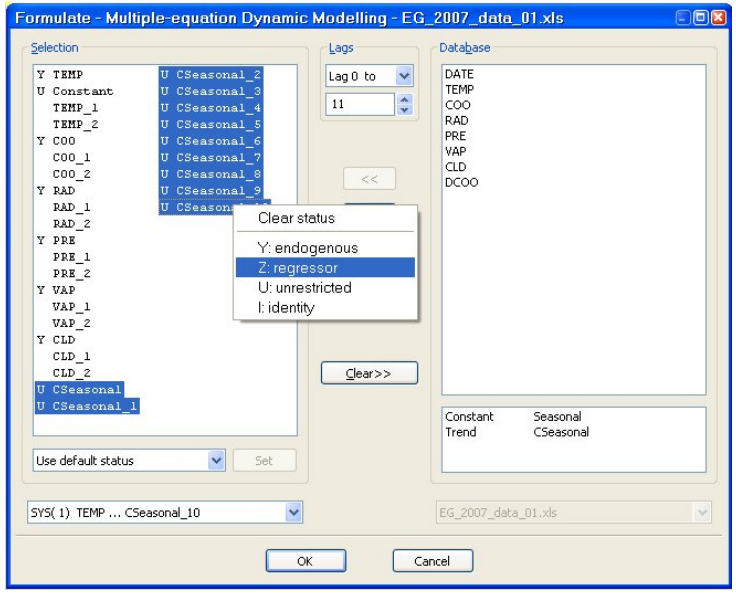
→ Model type: "Unrestricted system" (system of URFs), see Doornik and Hendry (2007b, chapter 3)

Formulate a system: (Model →) PcGive → Category: "Models for time series data" → Model class: "Multiple-equation dynamic modelling using PcGive" → "Formulate"



Multiple-equation modelling with Autometrics using second possibility. *Example*: 2007 Econometric Game, Question 2

- My GUM: A six-dimensional VAR(2) of $y_t = (TEMP_t, COO_t, RAD_t, PRE_t, VAP_t, CLD_t)$, with a constant and 11 centered seasonals in each of the six equations:



NOTE: Fixing variables (that is, not allowing Autometrics to delete them) now differs. Select the variables to delete → Right-click mouse → U: Unrestricted. (Unfixing: Z: regressor)

Results:

• NOTE: Autometrics simplifies even though the GUM does not pass all diagnostic checks

• Four variables are removed from all of the equations: The second lag of TEMP, VAP and CLD, and CSeasonal_10

Other type of analysis:

• Cointegration analysis (applied on the Unrestricted system, *not* on the simplified model): Model → Test → Dynamic Analysis and Cointegration Tests → ...

See Doornik and Hendry (2007b, chapter 4)

References:

Doornik, J. (2007). Autometrics. Working Paper, Economics Department, University of Oxford.

Doornik, J. A. and D. F. Hendry (2007a). *Empirical Econometric Modelling - PcGive 12: Volume I*. London: Timberlake Consultants Ltd.

Doornik, J. A. and D. F. Hendry (2007b). *Empirical Econometric Modelling - PcGive 12: Volume II*. London: Timberlake Consultants Ltd.

Hendry, D. F. and H.-M. Krolzig (2001). *Automatic Econometric Model Selection using PcGets*. London: Timberlake Consultants Press.

Hoover, K. D. and S. J. Perez (1999). Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. *Econometrics Journal 2*, 167–191.