

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. “ $D\text{COO} = \text{diff}(\text{COO},1);$ ” → 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):

The screenshot shows the OxMetrics software interface. The main window displays a data table with columns: DATE, TRHP, COO, RAD, FRK, VAP, and CLD. The 'DATE' column is highlighted, and the value '196401' is selected in the first row. A 'Change Sample' dialog box is open, showing the current database sample as '1 - 444 Undated'. The dialog box has the following settings:

- Current Database Sample: 1 - 444 Undated
- Frequency and Start Date:
 - Frequency: Monthly
 - Days per week: 5
 - Start Date: 1964-1
 - year-period
- Sample Size:
 - Observations: 0
 - Action: Add observations at the end

The dialog box has 'OK' and 'Cancel' buttons. The status bar at the bottom of the window shows 'DATE[1]' and '196401'.

	DATE	TRHP	COO	RAD	FRK	VAP	CLD
1	196401	-.46819	319.57	160.1	565.67	284.02	576.78
2	196402	-.63908	-99.99	169.3	529.84	279.12	561.82
3	196403	1.7239	-99.99	169.28	485.5	316.9	554.76
4	196404	-.86419	-99.99	170.36	561.06	258.69	542.07
5	196405	-.81742	322.23	166.66			
6	196406	-.11888	321.89	174.27			
7	196407	-1.1069	320.44	174.27			
8	196408	-.79489	318.7	169.99			
9	196409	-.84677	316.7	170.48			
10	196410	-1.8906	316.87	169.42			
11	196411	1.68	317.68	167.74			
12	196412	-.26573	318.71	169.88			
13	196501	-.67713	319.44	164.94			
14	196502	-.14814	320.44	173.85			
15	196503	-.90268	320.89	167.49			
16	196504	.014441	322.13	168.84			
17	196505	-.63035	322.16	172.3			
18	196506	1.8855	321.87	169.88			
19	196507	-.85229	321.21	169.41			
20	196508	-.57024	318.87	166.6			
21	196509	-.45379	317.81	169.38			
22	196510	-.90715	317.3	163.89			
23	196511	-.62274	318.87	169.44			
24	196512	-.32086	319.42	171.57			
25	196601	-.3568	320.62	163.62			
26	196602	-.79059	321.59	168.87			
27	196603	2.5034	322.39	170.66			
28	196604	-.28895	323.7	162.72	557.48	287.82	554.08
29	196605	-.50448	324.07	163.78	523.04	298.36	566.47
30	196606	1.6865	323.75	170.41	564.71	321.53	565.45
31	196607	-.28445	322.4	167.5	672.49	302.51	583.61
32	196608	-2.1055	320.37	170.17	735.03	248.43	570.47
33	196609	-1.9345	318.64	169.95	763.91	256.81	581.43
34	196610	-.17696	318.1	169.04	610.38	295.64	574.52

- *Example.* Create a differenced series:

The screenshot displays the OxMetrics software interface. The main window shows a data table with columns DATE, TEMP, COO, RAD, PRE, VAP, and CLD. The data spans from 1964(1) to 1966(10). The COO column contains several 'missing' values. An 'Algebra - EG_2007_data_01.xls' dialog box is open, showing the following code:

```
// Enter Algebra code here, for example:
Ly = log(y); Dly = diff(Ly, 1);

1 DCOO = diff(COO, 1);
```

The dialog box also includes a 'Functions' list with 'log(VAR);' selected, a 'Database' list with 'DATE', 'TEMP', 'COO', 'RAD', 'PRE', 'VAP', and 'CLD', and buttons for 'Run', 'Done', 'Load...', 'Save As...', 'Recal', and 'Write Algebra Code'.

	DATE	TEMP	COO	RAD	PRE	VAP	CLD
1964 (1)	196401	-1.46613	319.57	160.1	565.67	284.02	576.78
1964 (2)	196402	-1.63308	missing	163.3	629.84	273.12	561.82
1964 (3)	196403	1.7299	missing	169.28	485.5	316.9	554.76
1964 (4)	196404	-1.86419	missing	170.36	561.06	258.69	542.07
1964 (5)	196405	-1.81742	322.23	166.66	561.39	273.32	569.55
1964 (6)	196406	-1.1888	321.89	174.27	567.33	298.11	581.44
1964 (7)	196407	-1.1069	320.44				
1964 (8)	196408	-1.79489	318.7				
1964 (9)	196409	-1.84677	316.7				
1964 (10)	196410	-1.8906	316.87				
1964 (11)	196411	1.68	317.66				
1964 (12)	196412	-1.26573	318.71				
1965 (1)	196501	-1.67713	319.44				
1965 (2)	196502	-1.4814	320.44				
1965 (3)	196503	-1.90268	320.89				
1965 (4)	196504	-1.014441	322.1				
1965 (5)	196505	-1.63035	322.16				
1965 (6)	196506	1.8985	321.87				
1965 (7)	196507	-1.85229	321.21				
1965 (8)	196508	-1.57024	318.87				
1965 (9)	196509	-1.45379	317.81				
1965 (10)	196510	-1.90715	317.3				
1965 (11)	196511	-1.62274	318.87				
1965 (12)	196512	-1.32086	319.42				
1966 (1)	196601	-1.3568	320.62				
1966 (2)	196602	-1.79059	321.55				
1966 (3)	196603	2.5034	322.35				
1966 (4)	196604	-1.28895	323.7				
1966 (5)	196605	-1.50448	324.07				
1966 (6)	196606	1.6865	323.75				
1966 (7)	196607	-1.28445	322.4				
1966 (8)	196608	-2.1055	320.37				
1966 (9)	196609	-1.9345	318.64				
1966 (10)	196610	-1.17696	318.1				

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:
 - Estimation and inference while controlling for the impact from other variables: Spurious variables are more likely to be excluded, parameter estimates are “more correct”
 - 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:
 - Resource demanding and time consuming if properly implemented
- Solution: Automated Gets
 - 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

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

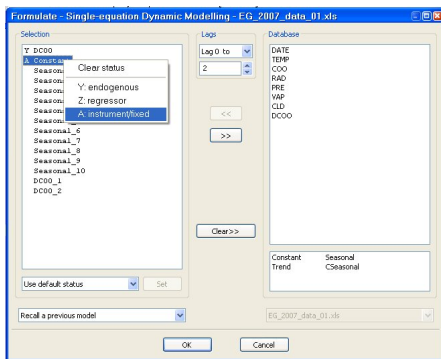
so a different specification is needed with, say, no lags of ΔCOO_t

- Formulate a model: (Model \rightarrow) PcGive \rightarrow Category: “Models for time series data” \rightarrow Model class: “Single-equation dynamic modelling using PcGive” \rightarrow “Formulate”

The screenshot shows the OxMetrics software interface. The background window displays a spreadsheet with columns for DATE, TEMP, COO, RAD, PPR, VAP, CLD, and DCOO. The DCOO column contains several 'missing' values. Overlaid on this is the 'PcGive - Models for time-series data' dialog box. The dialog box has a 'Module' dropdown set to 'PcGive' and a 'Category' dropdown set to 'Models for time-series data'. The 'Model class' dropdown is set to 'Single-equation Dynamic Modelling using PcGive'. At the bottom of the dialog, there are buttons for 'Formulate...', 'Estimate...', 'Test...', and 'Progress...'. The 'Formulate...' button is highlighted.

	DATE	TEMP	COO	RAD	PPR	VAP	CLD	DCOO
1964 (1)	196401	-.46813	319.57	160.1	565.67	284.02	576.78	missing
1964 (2)	196402	-.63308	missing	163.3	529.84	273.12	561.82	missing
1964 (3)	196403	1.7299	missing	169.28	485.5	316.9	554.76	missing
1964 (4)	196404	-.86419	missing	170.36	561.06	258.69	542.07	missing
1964 (5)	196405	-.81742	322.23	166.66	561.39	273.32	569.55	missing
1964 (6)	196406	-.11808	321.89	174.27	567.33	298.11	581.44	-.34
1964 (7)	196407	-1.1069	320.44	174.27	671.11	272.72	579.34	-1.45
1964 (8)	196408	-.79489	318.7	169.39	771.53	270.75	563.82	-1.74
1964 (9)	196409	-.84677	316.7	170.48	768.29	277.99	579.39	-.2
1964 (10)	196410	-1.8906	316.87	169.42	659.14	262.62	590.88	.17
1964 (11)	196411	1.68	317.68	167.74	596.23	328.01	578.7	.81
1964 (12)	196412	-.26579	318					
1965 (1)	196501	-.67713	319					
1965 (2)	196502	-.14814	320					
1965 (3)	196503	-.90268	320					
1965 (4)	196504	.014441	322					
1965 (5)	196505	-.63035	322					
1965 (6)	196506	1.8855	321					
1965 (7)	196507	-.85229	321					
1965 (8)	196508	-.57024	318					
1965 (9)	196509	-.45379	317					
1965 (10)	196510	-.90715	31					
1965 (11)	196511	-.62274	318					
1965 (12)	196512	-.92086	319					
1966 (1)	196601	-.3568	320					
1966 (2)	196602	-.79059	321					
1966 (3)	196603	2.5034	322					
1966 (4)	196604	-.28895	32					
1966 (5)	196605	-.50448	324					
1966 (6)	196606	1.6865	323					
1966 (7)	196607	-.28445	32					
1966 (8)	196608	-2.1055	320					
1966 (9)	196609	-1.9345	318.64	169.95	762.91	266.82	581.43	-.73
1966 (10)	196610	-.17696	318.1	169.04	610.38	295.64	574.52	-.54

- 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:

Autometrics Settings - Single-equation Dynamic Modelling

<input type="checkbox"/> Search settings	
Outlier detection	None
Pre-search lag reduction	<input type="checkbox"/>
Pre-search variable reduction	<input type="checkbox"/>
Search effort	1
Backtesting	GUM 0
Tie-breaker	SC
Print level	Default output
Target size	Default: 0.05
User determined p-value	.05
Diagnostic test p-value	.01
Standard errors	Default
GIVE: first do reduced form	<input checked="" type="checkbox"/>
<input type="checkbox"/> Block identification when there are too many parameters	
<input type="checkbox"/> Diagnostic test set	
Use default	<input checked="" type="checkbox"/>
Normality test	<input checked="" type="checkbox"/>
Heteroscedasticity test (using squares)	<input checked="" type="checkbox"/>
Heteroscedasticity test (using squares and cross products)	<input type="checkbox"/>
Chow test	<input checked="" type="checkbox"/>
RESET test (using squares)	<input type="checkbox"/>
Error autocorrelation test	<input checked="" type="checkbox"/>
Portmanbeau statistic	<input type="checkbox"/>
ARCH test	<input checked="" type="checkbox"/>
<input type="checkbox"/> Diagnostic test arguments	
Use default	<input checked="" type="checkbox"/>
Chow-test sample split (%)	70
Error autocorrelation to lag	2
Portmanbeau lag length	10
ARCH test to lag	2

OK Cancel

- 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]

```

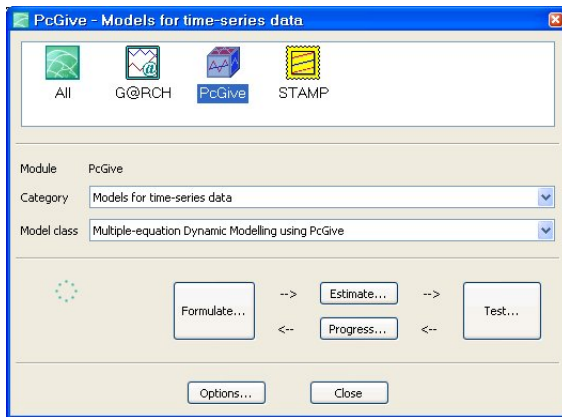
Further diagnostic tests:

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

Multiple-equation modelling with Autometrics: Two possibilities

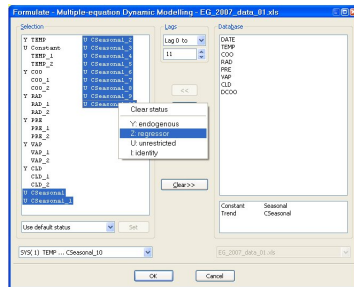
- Seemingly Unrelated Regression (SUR) using OLS/IV, that is, single-equation modelling of each equation separately (requires stationarity of regressors)
- Simultaneous variable deletion (or non-deletion) across equations using vector diagnostic tests but estimation still by OLS (does not require stationarity 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 \rightarrow) PcGive \rightarrow Category: “Models for time series data” \rightarrow Model class: “Multiple-equation dynamic modelling using PcGive” \rightarrow “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 \rightarrow Test \rightarrow Dynamic Analysis and Cointegration Tests \rightarrow ...

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.