UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Master's
Applied Time-Series Econometrics

Session 5: Building a system of equations model to conduct macroeconomic forecasts

Lecturer:
Nico Katzke
nicokatzke@sun.ac.za

Department of Economics

DEPARTMENT OF ECONOMICS

Plan

- In this session we look at developing a systems of equations (SOE) model.
- The main focus will be on the **practical sessions** and getting our modelling right there.
- This session will highlight some of the key issues that are unique to this method of modelling.

Multivariate models

- Today we build on the Multivariate VAR and VECM by extending our analysis to the dynamic multivariate framework using a systems approach
  - In particular, we also include and model exogenous variables (and their lags) to explain the behaviour of our target variables, which themselves interact dynamically in the system.
- This is then followed by setting up several equations that are linked to one another through a SOE-model.

Reduced form and structural equations

- Standard macroeconomic theory often requires estimating models that constitute an entire system of equations. Consider e.g. Samuelson's classic (1939) model of output determination (Keynesian model):

$$y_t \equiv c_t + i_t + \mathrm{g} \qquad (1)$$

$$c_t = \alpha(y_{t-1}) + \varepsilon_{c,t} \qquad (2)$$

$$i_t = \beta(c_t - c_{t-1}) + \varepsilon_{i,t} \qquad (3)$$

$$y_t \rightarrow output, \quad c_t \rightarrow consumption, \quad i_t \rightarrow Investment, \quad g \rightarrow gvt\ spending$$

- From this it is clear that $\boldsymbol{y_t}, \boldsymbol{i_t}$ & $\boldsymbol{c_t}$ are **endogenous** variables in the system (in that it is explained by variables in the model), and **g** is **exogenous** (and we have no equation to explain this…)
- Equation 1 = an **identity** (must hold).
- Eq 2 & 3 are **endogenous** equations for c_t & i_t, while $g_t \rightarrow$ exogenous variable
- y_{t-1} & c_{t-1} → **Lagged** (or predetermined) endogenous variables
- $\varepsilon_{c,t}$ & $\varepsilon_{i,t}$ → zero mean random disturbances (i.e. unexplained part)

Reduced form and structural equations

- From the previous slide, Equation (3) → is a **structural equation**, as it expresses the (here endogenous) variable i_t as a function of the **current realization** of **another** (also endogenous) **variable** c_t.

- **A reduced-form equation** represents a variable as a function of its **own lagged** values, **lag values** of other endogenous variables and / or **current values** of ***exogenous*** variables and disturbance terms. It **does not include contemporary** (or current) **values** of any **endogenous** variable.

Reduced form and structural equations

- In the model then, Investment is **NOT** in reduced form, as it depends on the **current level** of consumption, which is an **endogenous** variable in the system of equations.
- We can easily rewrite **Investment** in **reduced form** by substituting (2) in (1) to yield:

$$i_t = \beta\left([\alpha y_{t-1} + \varepsilon_{c,t}] - c_{t-1}\right) + \varepsilon_{i,t}$$

Thus: $$i_t = \alpha\beta y_{t-1} + \beta c_{t-1} + \varepsilon_{i,t} + \beta\varepsilon_{c,t-1}$$

Which is now in reduced form.

(This form is, however, not unique → as we could further sub in equation 2's lag for c_{t-1}, etc.)

Reduced form and structural equations

2. <u>**Reduced form & structural equations:**</u>

Similarly, we could rewrite the output equation by substituting (2) and (3) into equation (1):

$$y_t = c_t + i_t = [\alpha y_{t-1}] + [\beta(c_t - c_{t-1})] + \varepsilon_{i,t} + \varepsilon_{c,t}$$

$$= \alpha y_{t-1} + \beta\big([\alpha y_{t-1} + \varepsilon_{c,t}] - c_{t-1}\big) + \varepsilon_{i,t} + \varepsilon_{c,t}$$

$$= \alpha y_{t-1} + \beta\big([\alpha y_{t-1} + \varepsilon_{c,t}] - [\alpha y_{t-2} + \varepsilon_{c,t-1}]\big) + \varepsilon_{i,t} + \varepsilon_{c,t}$$

$$= \alpha(1+\beta)y_{t-1} + \alpha\,\beta(y_{t-2}) + \alpha(1+\beta)\varepsilon_{c,t} + \beta\varepsilon_{c,t-1} + \varepsilon_{i,t}$$

Which is now a **univariate** reduced form equation of y_t, as it is a function only of its **own lagged** values and disturbances: Thus it now is effectively an ARMA(2,2) process → $y_t = f(y_{t-1}, y_{t-2}) + f(\varepsilon_{c,t-1}, \varepsilon_{c,t}, \varepsilon_{i,t})$

Problem of simultaneity

- Consider again the system of equations for the Keynesian output function.
- Suppose we used **OLS** to run the regression for the model specified – surely it would lead to biased results. That is because the equations are contemporaneously <u>**related**</u>
  - Eq 1 contains the contemporary values for consumption and investment – which is in turn specified by their own endogenous equations…
  - Clearly, ordinary OLS techniques would not suffice – as these equations would yield biased and non-sensical results due to the problem of simultaneity bias
- This is because there is a feedback effect – as c explains y, y explains c and Δc explains i.

System of Equations

- By re-specifying the equations to be in structural form, we can run the model if it is **identified**.

- Identification implies there is enough information to enable the structural form coefficients to be estimated from its reduced form

- We omit a further discussion on identification issues as it requires advanced understanding of matrix algebra falling outside the practical nature of this course.

Why use this?

- What follows is an intuitive explanation of what we will be covering in the practical classes regarding the fit of models that forecast a variable based on defining its sub-parts in the system (like the Keynesian model we specified earlier).
- This is especially useful when setting up models that forecast based on theoretical specifications.
- It allows variables to effect each-other in a dynamic setting

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

System of equations approach

The system of equations approach to modelling macroeconomic relationships

- Let's consider a theoretic justification for determining GDP output.
- In particular, consider the following more detailed theoretic framework:
  - Note: $x_t = \mathbf{log}(X_t)$ for all the cases below.

Macroeconomic Theory states:

$$\boldsymbol{GDP = Cons + Inv + G + EX - Im}$$

$$Cons = f\big(GDP, cons(-1)\big)$$

$$Inv = f\big(GDP, \Delta GDP(-1), r(-4)\big)$$

$$r = f\big(GDP, \Delta GDP, \Delta(money\ supply), .$$

Or in equation form:

$$y_t = c_t + i_t + g_t + EX_t - IM_t$$

$$c_t = \alpha_1 + \beta_1 . y_{t-1} + \beta_2 . c_{t-1}$$

$$i_t = \alpha_2 + \beta_3 . (\Delta y_{t-1}) + \beta_4 . c_t + \beta_5 . r_{t-4}$$

$$r_t = \alpha_3 + \beta_6 . y_t + \beta_7(\Delta y_t) + \beta_8(\Delta m_t) + \beta_9(\Delta r_{t-1})$$

Reduced Form

- As can be noted, i_t is again a **structural equation**, as it expresses the endogenous variables (i_t) as a function of the current realization of **another** (also endogenous) **variables** y_t.
- Thus Investment is NOT in reduced form, as it depends on the **current level** of consumption, which is an **endogenous** variable in the system of equations.
- But again, we can put **Investment** in **reduced form** by substituting (1 & 2) in (3).
- This would allow us to estimate parameters for the equation.
    - Note: if this is possible, Eviews will do so automatically and does not require the modeller to do so by hand.

The system of equations approach to modelling macroeconomic relationships

- En Endogenous variables: series that have equations explaining their behaviour inside the model.
  - For it to be endogenous, we need to be able to write it as: $Y = f(x, y, z)$
- X Exogenous variables: Having values determined outside the model
- A Add-factors: special case of exogenous variables returned to later.

- NB to remember that when we forecast – we need to specify values for **exogenous** variables ourselves (as it is not explained within the model…)
  - Typically we will use univariate autoregressive techniques in providing simplified forecasts for it, or use **qualitative information** (e.g. we believe G to increase by 5% for the next year…)

The system of equations approach to modelling macroeconomic relationships

- Identities: these are equations that we expect to hold exactly. Thus the $GDP = C + I + G$ equation would be an **identity**. These equations have no parameter values and are assumed to hold by definition.
- Stochastic equations: We expect these equations to have random errors and are not perfectly specified (like the cons, inv and r equations).
- Typically we first fit the **stochastic equations** to approximate the parameters and check for approximate stationarity. After we are happy with our equation fit, we can build the model to forecast.

Simultaneity

- The problem of simultaneity, as mentioned before, can bias our parameter estimates, and requires a **simultaneous fitting** of the equations. In this way, we can avoid, e.g., GDP being correlated with the residuals of other equations in which it enters contemporaneously, and thereby violating OLS assumptions.
- For this, we fit a dynamic system of equations.
  - The process we follow in the tuts might seem a bit tedious, but it is robust and teaches us how to use the system estimation.

Steps to follow

- We will use the following logical progression to fitting our theoretic Keynesian cross model:

1. Fit the best endogenous equations (with OLS) as suggested by theory. This involves checking the fit and conducting necessary checks to **ensure the best possible model explaining the series**.
2. Thereafter, run the equations simultaneously using a system (e.g. FIML) approach.
3. Fit a model with the equations specified in 1 and fitted in 2.
4. Specify the identity used.
5. Run the model and update the links.
6. Evaluate the forecasting abilities of the model
7. Use the model to forecast.

Step 1: fitting the equations

- In this step we estimate the fit of the equations proposed in theory.
- Apply the knowledge that we have gained so far from univariate time-series, in that we need to first remove a unit-root if present, and also take the natural log where needed.
    - NB – we also need to control for long run effects (cointegrating equations) where applicable.
- Thereafter we include other explanatory variables where needed.

Step 1: fitting the equations

- Remember our goal is to fit several endogenous equations into a model to explain the dynamics of the identity:

$$Y = C + I + G + EX - IM$$

- And then use it to forecast…
- For this reason, our individual endogenous equations need to control for serial autocorrelation, but also include enough information to allow the equations to affect each other.
- Thus we don't fit ARIMA factors as explanatory now for the sake of simplicity. Rather stick to using only exogenous factors to explain the behaviour…

Issue of errors-in-variables

- When working with broad macroeconomic data as we are, we often run into the problem of errors-in-variables.
  - This implies that the variables in a regression is measured with random error
  - Errors-in-variables then leads to biased coefficients in ordinary least squares. Sometimes this can be fixed with 2SLS.
- As our data naturally contains errors and often these errors could be correlated with other included variables, we should consider using 2SLS if our coefficients do not make sense.

Issue of errors-in-variables

- Instead of 2SLS, econometricians today are more likely to use the *Generalized Method of Moments*, or *GMM*.
- Two-stage least squares can be thought of as a special case of GMM. GMM extends 2SLS in two dimensions:
  - GMM estimation typically accounts for heteroskedasticity and/or serial correlation.
  - GMM specification is based on an orthogonality condition between a (possibly nonlinear) function and instruments.
- In Eviews, we can type such an equation out as:

gmm Y c X1 X2 @ c X1(-1) X2(-4)

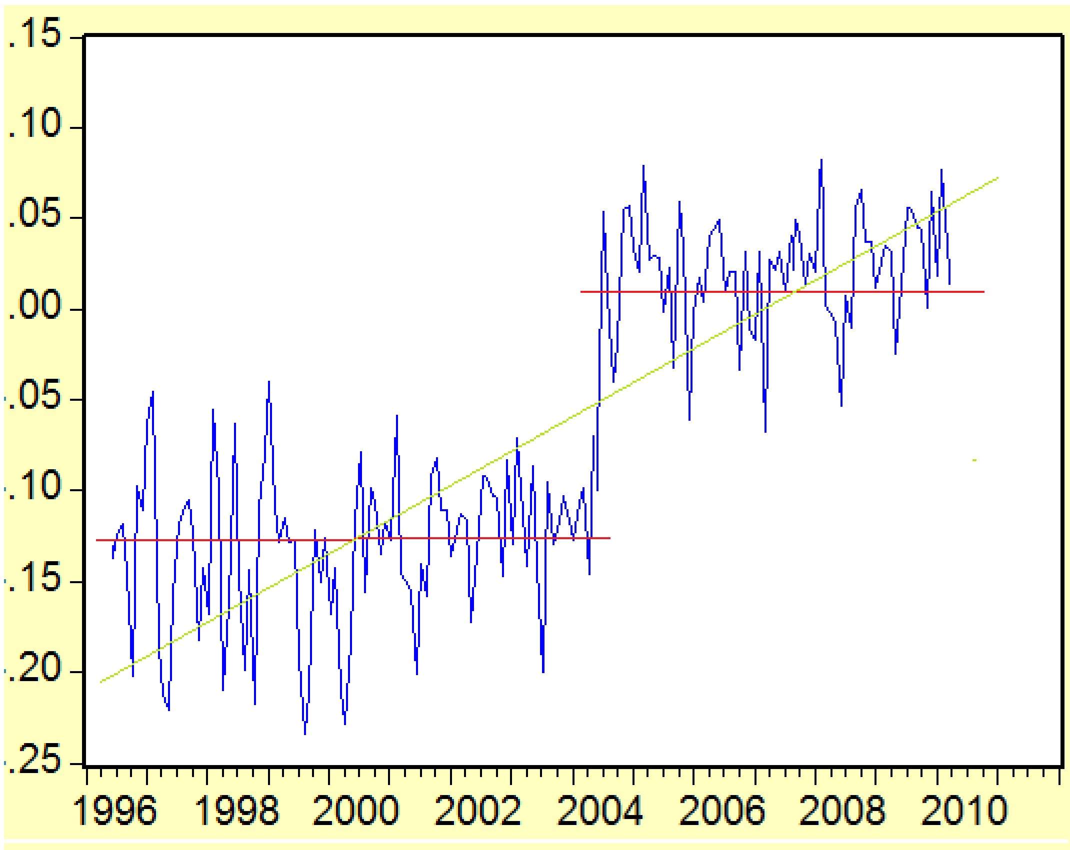
Where the instrumental variables are, in this example, lags of the series included.

Controlling for significant events

Controlling for structural changes

- Before proceeding, let's first consider significant events…
- As done in the first quarter, we can test formally for the presence of structural breaks using a **Chow test** at a specific date.
- But how important is it to control for structural breaks in modelling time series variables?
- Could we not simply ignore such periods?
  - Firstly, a significant event could bias the coefficients that are used to explain relationships and to forecast values.
  - Significant events could then lead to significant tests being rejected due to one or two isolated events.
  - This may be especially valid if a series seems stationary before and after an event, but is constant at a different level before and after such an event. This is known as a structural change!

Structural change in the series



- Illustrating this latter point, suppose the series is roughly stationary (red line) up till 2004, when there is a **sudden regime change.**
- Thereafter y_t experiences a significant one-time jump and is again roughly stationary thereafter.
- The green line would be the DF test used if we do not account for this regime change, possibly leading to the non-rejection of a unit root process, when the series is actually still stationary after the event… but only at a higher level.
- This would also lead to an upward bias of the coefficients in this case.

Importance of controlling for structural breaks in series

- From the figure on the previous slide it is clear that not controlling for the structural change will yield an inaccurate OLS estimation (green line).
- If we include a dummy variable having all values equal to zero before and all values equal to one thereafter, it allows the intercept to "jump" permanently and thus take into account the structural change (with the series again possibly being stationary thereafter!)
  - ***Note NB that this is different from a one-time significant event that sees the series return to its previous level immediately thereafter!!

Difference between a shock and a structural break

- Note that a **structural break** is different from a **one-time significant event** that sees the series return to its previous level immediately thereafter…
- A **structural change** sees a permanent change in the level of the series thereafter. Not controlling for a one-time event may not necessarily invalidate the DF unit root test (although it may bias the coefficients), but a structural change more likely would!
- **<u>Controlling for such events:</u>**
  - Controlling for a **one-time significant** event : we may specify a dummy variable that equals one at the date, and zero otherwise.
  - If we want to control for a structural change, we include a dummy variable having all values equal to zero before and **all values** equal to one thereafter. This allows the slope to "jump" permanently and thus take into account the structural change.

Using such tests

- We will use the Chow Break-Point test in the practical classes as a way of assessing whether we need to account for structural breaks in our equations.
- We could merely use the statistical significance test to assess whether one-time events should be controlled for using a dummy.

- The question then is should we use an abrupt structural break (0 before t, 1 thereafter), or should the transition be gradual? As suggested in Ruschelrath's paper on competition policies in Germany…

Controlling for a break-point… Ruschelrath

- As can be seen on the right, it might at times make more sense for the indicator variable to go from $0 \rightarrow 1$ at a **gradual pace**, as opposed to an abrupt change…
- This depends on the type of change!

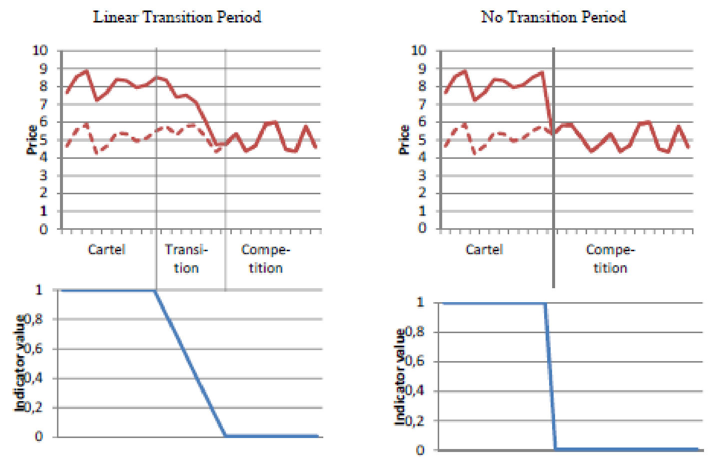


Figure 3: Design of the Indicator Variable ‚Cartel Period'

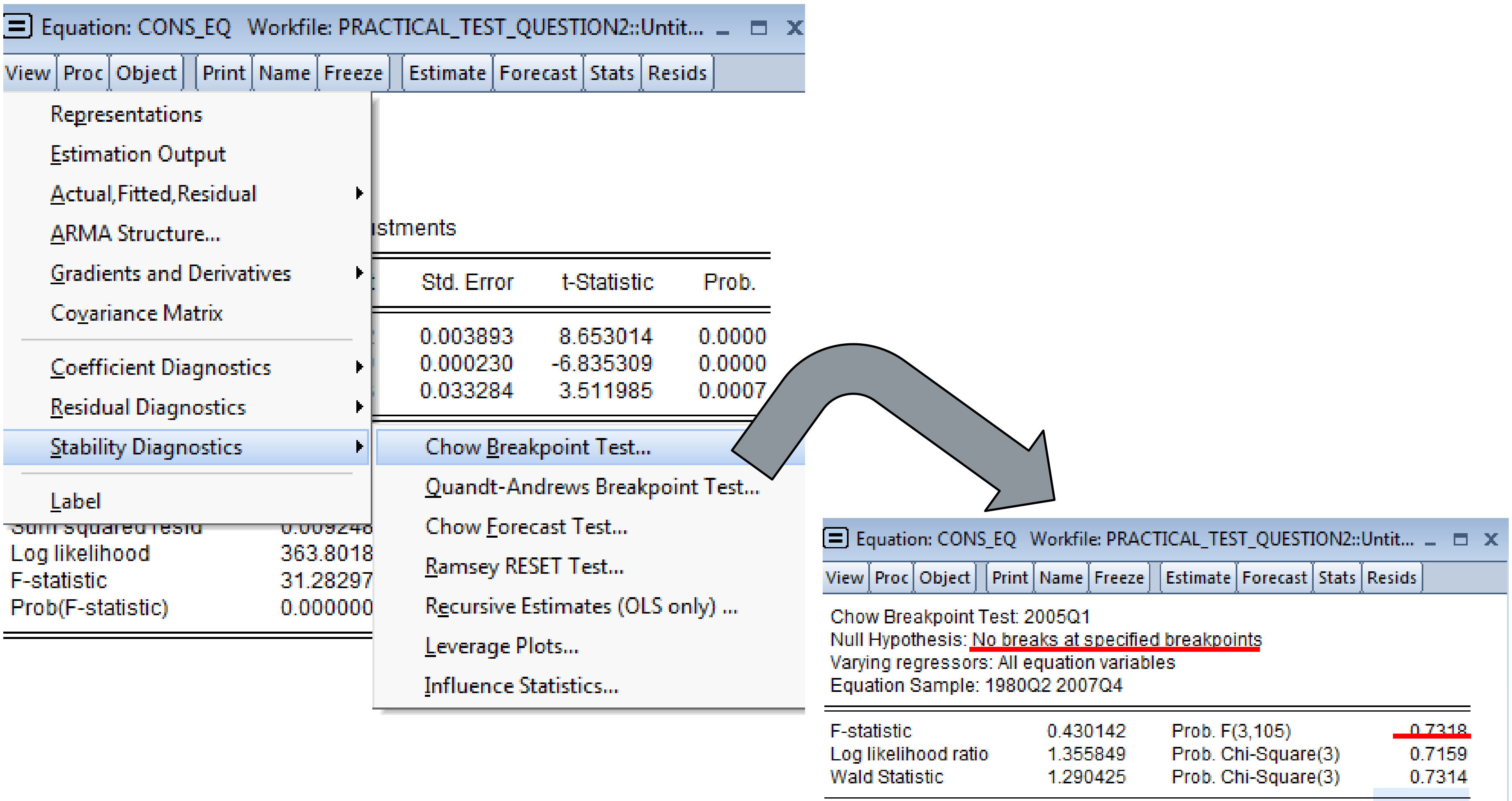
Source: Own Figure

Instead of an introduction of a continuous indicator variable, it would alternatively be possible to introduce an additional indicator variable which has a value of '1' in the transition period and a value of '0' otherwise. Although such a solution does not assume a specific functional form, it still allows accounting of the transition period.

Based on this brief general discussion of the relevance of considering the transition, we continue by comparing the effect of different transition periods on the price overcharge estimate. In the following, we apply the extended econometric model (following the instrument variable approach) and include the transition period in the estimation of the price overcharge. In particular, we differentiate between three transition periods: January to April 2002 (columns (1) and (2) in Table 4), January to June 2002 (columns (3) and (4)) and January to July (columns (5) and (6)). As reference point, column (0) in Table 4 repeats the results of the estimation of the extended econometric model above. Models (1), (3) and (5) use an indicator variable to mark the transition period (which is treated as part of the cartel period). In models (2), (4) and (6), a continuous transition is assumed, i.e. in column (2), the indicator variable has a value of 0.8 in January, a value of 0.6 in February and so on.

Using the Chow Test

- We use the Chow test as follows:



Moving on to step 2…

- After fitting equations for all the endogenous factors and controlling for significant events / structural breaks, we employ a **simultaneous fitting** of the equations to overcome the simultaneity bias mentioned earlier.
- We do this by copying the equations into a system…
  - Then we set up the system by typing *system* at the top.
  - You will then have a blank screen, in which you type the equations (in the format specified earlier).
  - By employing a simultaneous equation procedure – we avoid the simultaneity bias that would have made the results largely invalid due to endogenous variables in the equations…

Simultaneous equations

- There are then many techniques that we can use to solve our equations simultaneously.
- One would be to use the **SURs** (Seemingly Unrelated Regression) approach – where we allow for the contemporaneous relationships between the error terms in the equations. Thus under the SUR framework the error terms are transformed so that they become unrelated. It, however, requires many parameter estimations…
- Other approaches include the **2SLS**, **IV** and **GMM** approach.

- As we do not specify any instrumental variables, we rather use the **FIML** approach – which is a Full-Information Maximum Likelihood approach.

FIML

- Without going into too much detail (most statistical textbooks discuss the techniques in greater depth), the FIML approach considers all the parameters jointly – and an appropriate likelihood function is then maximized.

- The assumption is that the residuals have a joint normal distribution.
  - Using this method, we have now established values for the parameters that are efficient, and not affected by the simultaneity bias that would have affected its fit had we not used a simultaneous equations approach.
  - *This approach should be largely equivalent to the SUR approach

Parameter values…

- After hitting estimate, you now have the parameter estimates of a system that has been fitted simultaneously.
- These parameter estimates are now fully efficient and the implicit endogeneity problem controlled for.

S System: SYS01 Workfile: C:\USERS\NICOKATZKE\DESKTOP\DESKTOP FOLDE...

View | Proc | Object | Print | Name | Freeze | InsertTxt | Estimate | Spec | Stats | Resids

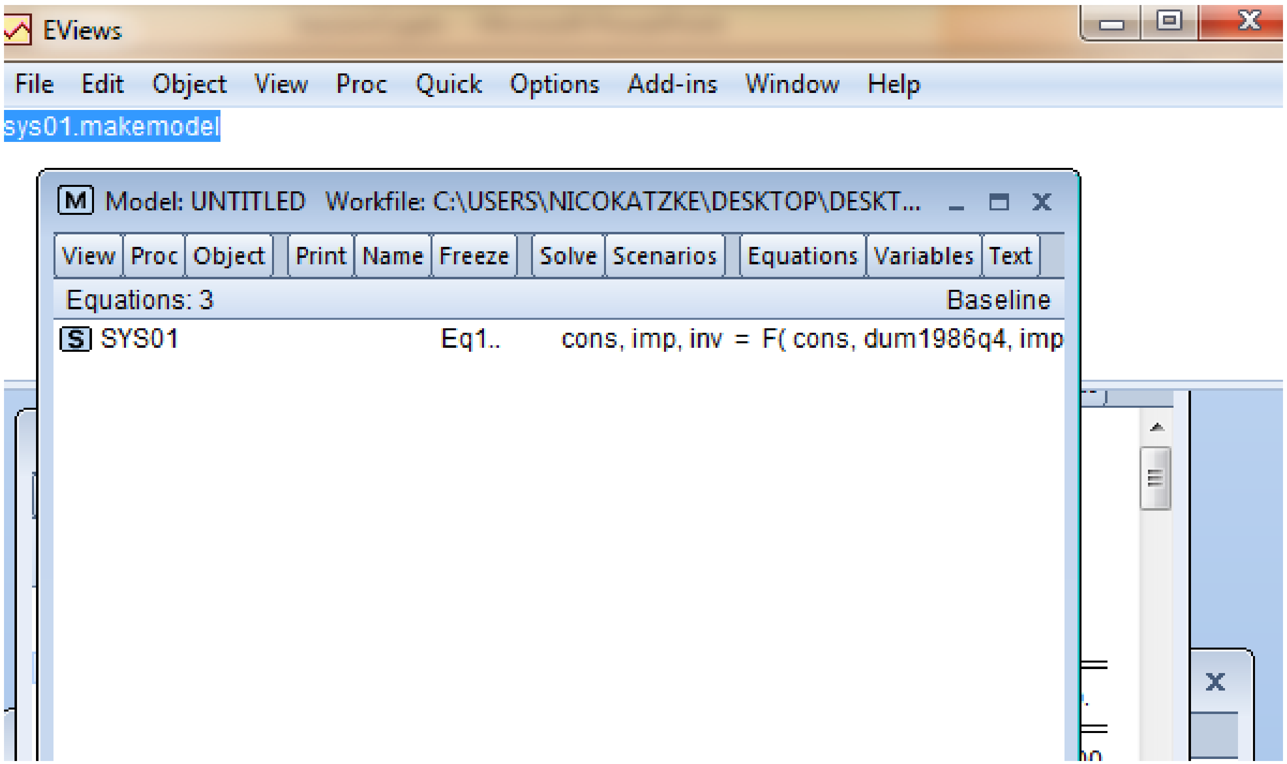
System: SYS01
Estimation Method: Full Information Maximum Likelihood (Marquardt)
Date: 04/16/13 Time: 18:06
Sample: 1980Q3 2007Q4
Included observations: 110
Total system (balanced) observations 330
Failure to improve Likelihood after 69 iterations

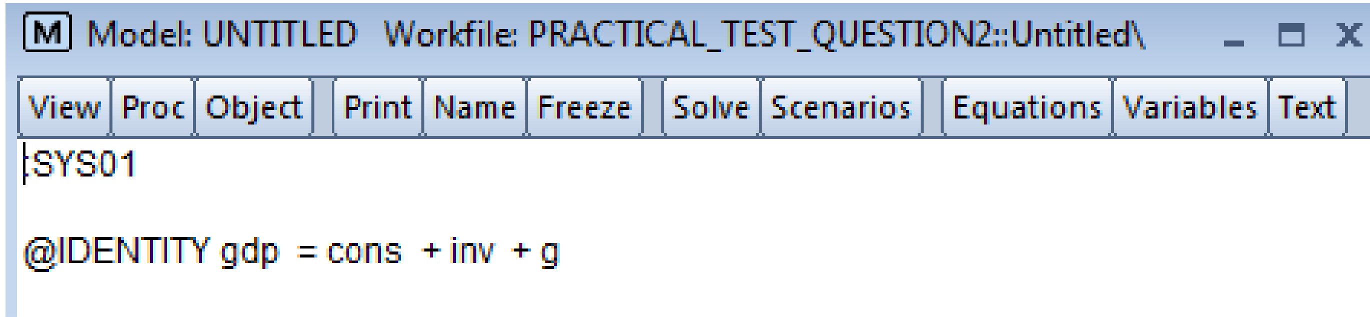
	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.030975	0.004984	6.214303	0.0000
C(2)	-0.001426	0.000289	-4.930563	0.0000
C(3)	0.140497	0.028706	4.894244	0.0000
C(4)	0.001450	0.006024	0.240649	0.8098
C(5)	2.381861	0.692540	3.439312	0.0006
C(6)	-0.114767	0.080827	-1.419901	0.1556
C(7)	-0.274988	2.540080	-0.108259	0.9138
C(8)	0.055056	0.021333	2.580843	0.0099
C(9)	1.068636	0.525034	2.035366	0.0418
C(10)	-0.003078	0.001212	-2.540067	0.0111

Log likelihood	-976.2260	Schwarz criterion	18.17688
Avg. log likelihood	-2.958261	Hannan-Quinn criter.	18.03096
Akaike info criterion	17.93138		
Determinant residual covariance	1.53E-10		

Making the model

- After we have successfully fitted the FIML simultaneous equations system, we name it (sys01), and then type at the top: sys01. makemodel. It should look as follows:



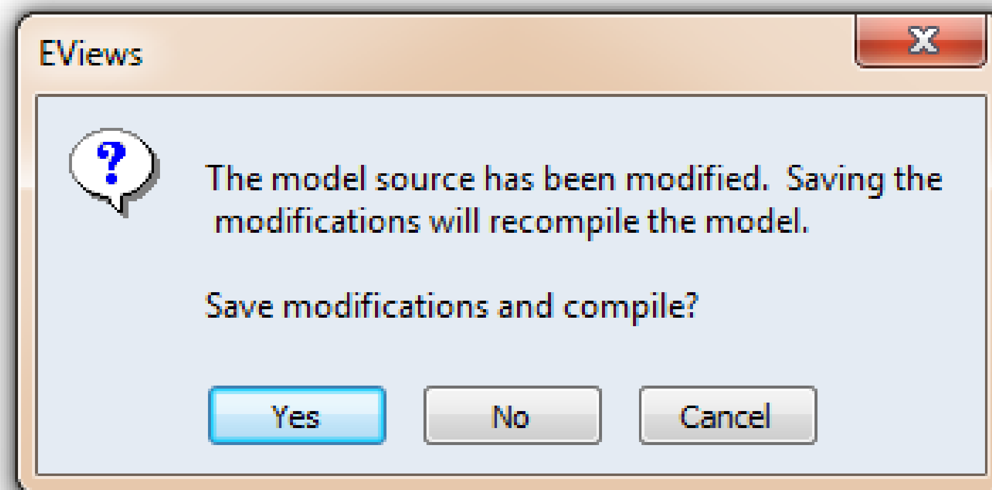


In the model window – select TEXT.

And then type the **identity**:

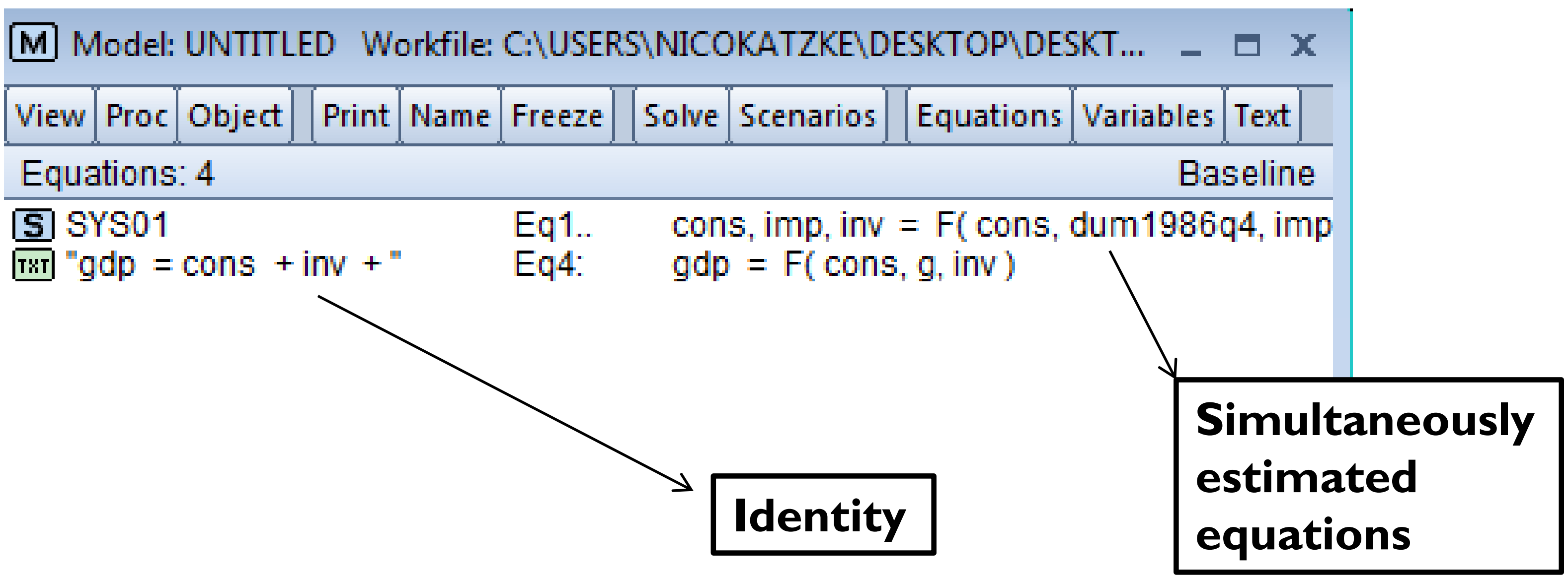
@identity Gdp = cons + inv + g

Then select Equations at the top – you will then be prompted to modify the model source.
Select yes.



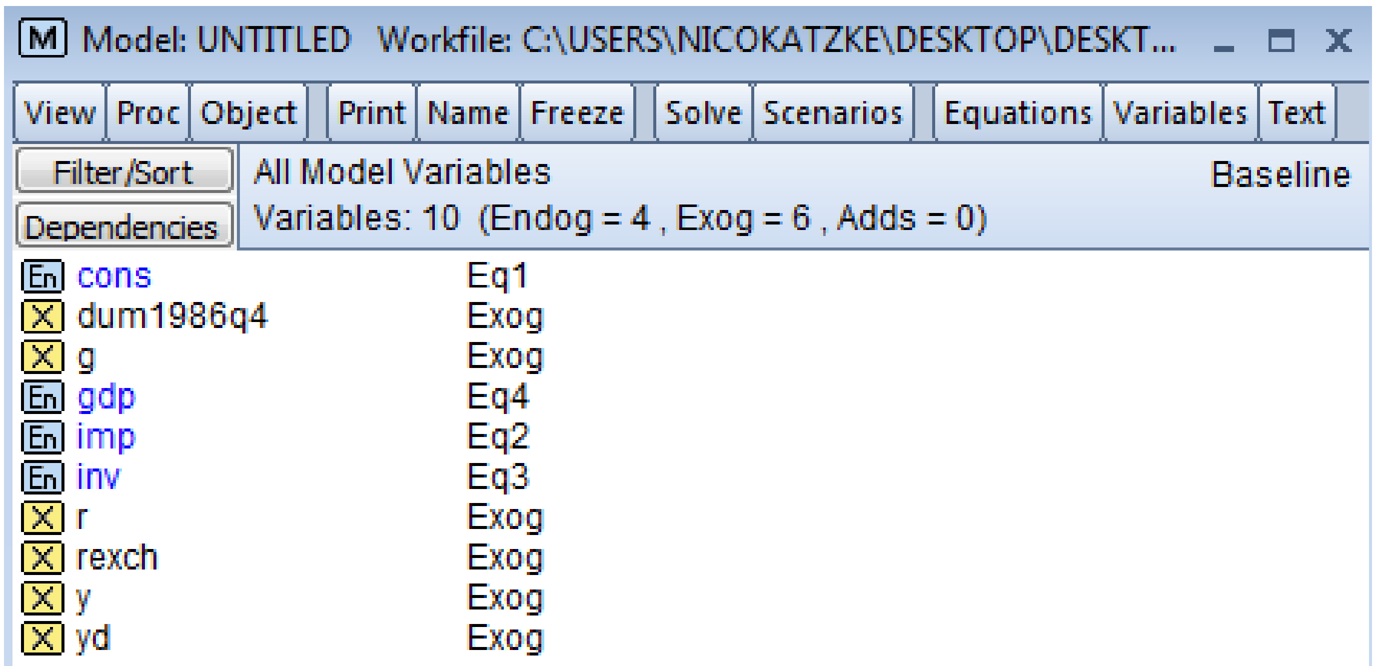
Making the model

- After selecting Equations you will see:



Making the model

- Selecting Variables, you will now see it as Endogenous / Exogenous



The system of equations approach to modelling macroeconomic relationships

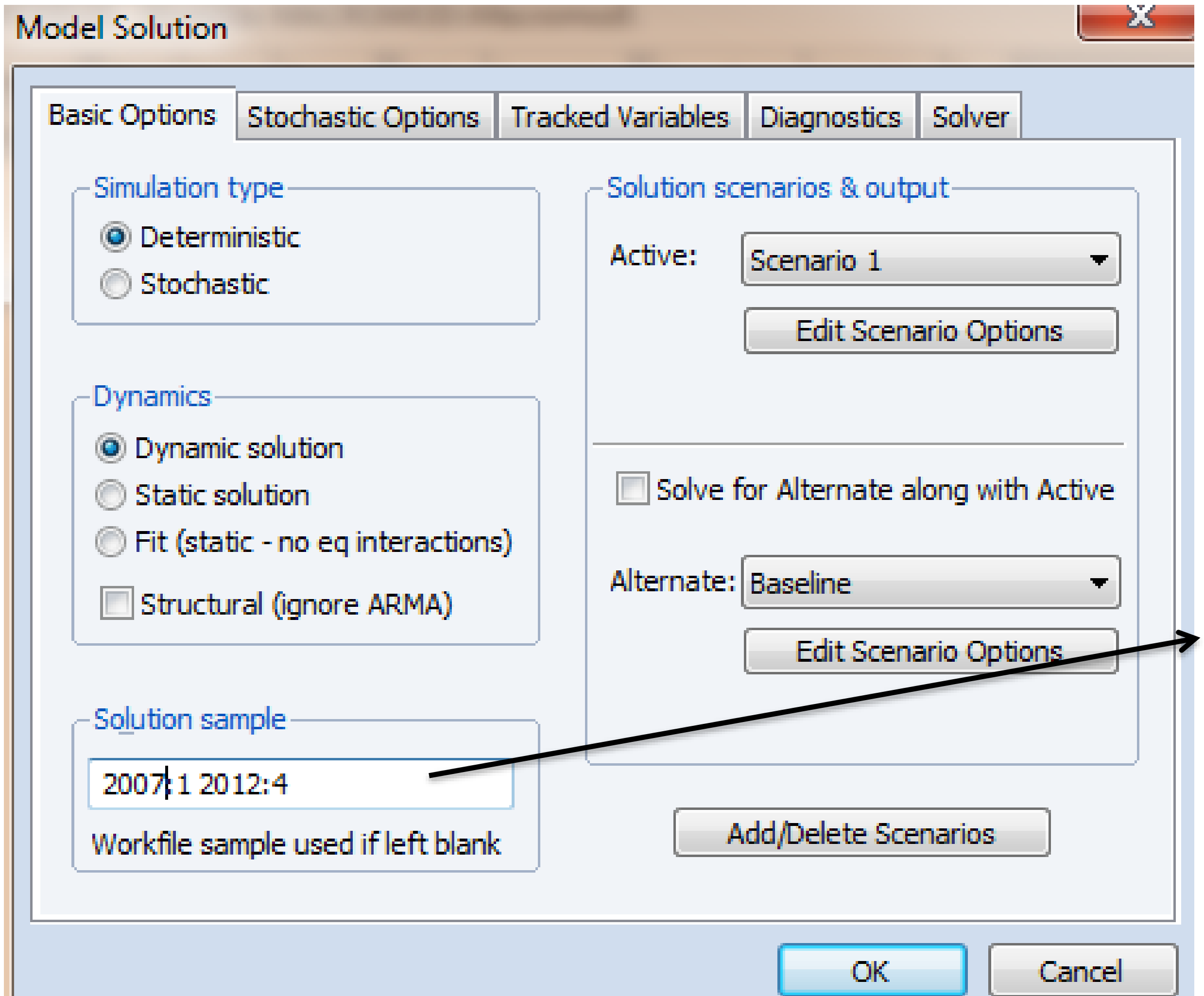
- After specifying the model, we want to run it to see **how well** our intended model can **actually explain values** of our endogenous variables for which we have equations.
  - Remember that we can either conduct **static** in-sample forecasts – where we use the **true** values for the explanatory variables in the stochastic equations (e.g. in the equation: $c_t = \alpha_1 + \beta_1 . y_{t-1} + \beta_2 . c_{t-1}$ we use the true historic values of y_{t-1} & c_{t-1} when fitting model estimates for c_t).
  - Or we can conduct the more realistic approach of **dynamic forecasting** – where we test how well we would have forecast c_t into the future, not knowing the true values of the variables in the equation (e.g. estimating c_{1995} using $\widehat{y_{1994}} \& \widehat{c_{1994}}$ that we estimated using 1993 data)…

The system of equations approach to modelling macroeconomic relationships

- **Solving** implies Eviews runs the system of equation and fits values to the endogenously specified variables.
- It gives us aliases (or nicknames) to the forecasted variables (such as y_0) with which to compare the true values of the variables (y).
- Note that this now allows us to evaluate the in-sample forecasting ability of our model!

The system of equations approach to modelling macroeconomic relationships

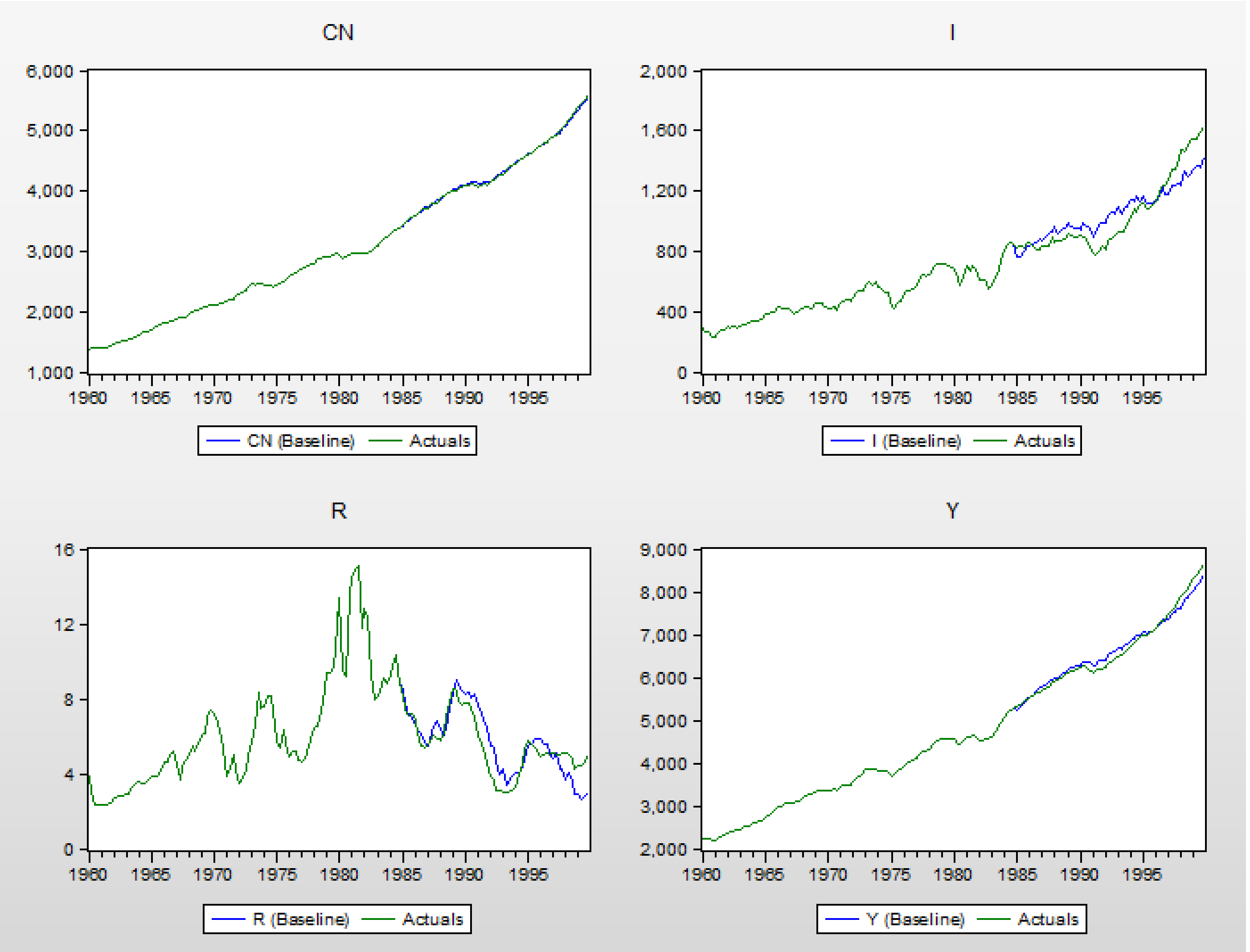
- Selecting **Solve** will bring up the following screen:



For in-sample testing, remember to specify a time period within our sample…

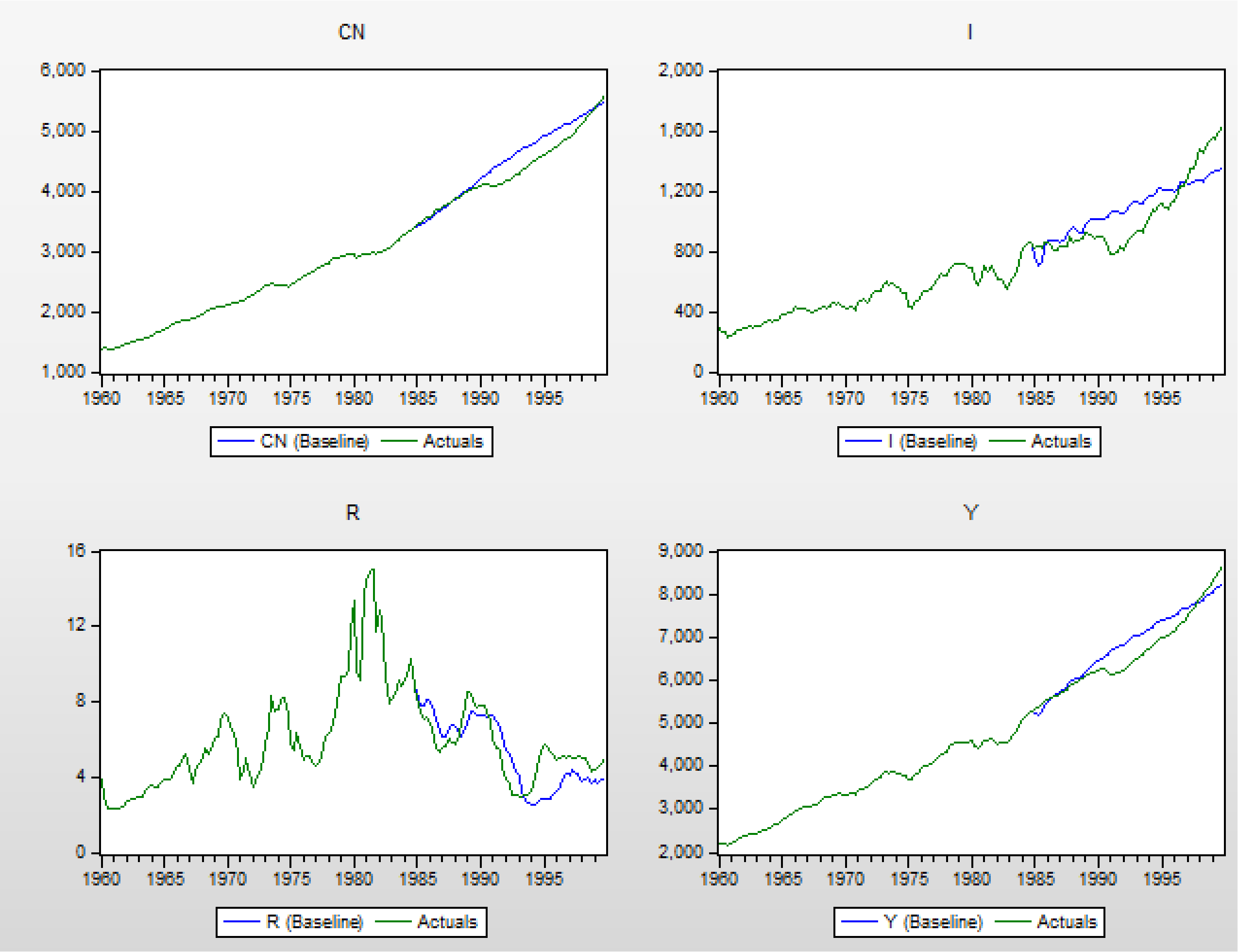
The system of equations approach to modelling macroeconomic relationships

- Suppose we did a **static in-sample test** on the period: 1985 – 1990.
- Here are graphs of the endogenous variables for a one-step ahead static forecast approach.
- (Why only graph the endogenous variables? Do the exogenous variables change?)



The system of equations approach to modelling macroeconomic relationships

- The **dynamic** forecasts clearly do not provide us with a similarly tight fit…
- This follows as it uses estimates for the variables included in the equations, and not the actual values…
- This gives us a better indication of the model's accuracy of forecasting into the future…



The system of equations approach to modelling macroeconomic relationships

- So… we can use a ***static forecast*** just to check whether our equations and parameters are any good at tracking the data well if we knew the inputs… and then we use a **dynamic forecast** to see how well we would have forecast the period: 1985 – 1990 if we were at 1984 and had to use estimated endogenous parameters in our models…
- BUT (beeeg but): how were the exogenous variables specified in our dynamic in-sample test (remember they have no equations that estimate their values – like government spending g_t)?!!
  - **For a dynamic in-sample test – we use the true exogenous values.**

- But when we forecast into the future – of course we don't know the **true exogenous values** which are used in forecasting (in addition to the endogenous variables) the stochastic equations… (what G going to be in 2014?!)

The system of equations approach to modelling macroeconomic relationships

- So, after we have checked that our model is somewhat capable of conducting forecasts, the next step would be to "create" values (or guestimates) of what the exogenous values would be for, say, 2014 – 2020.
- We will look at a variety of ways to tackle this issue – but it differs for the type of variable.
  - Economists might have a good explanation of where money supply might head to, but less of an idea what government spending might do in the next few years.

The system of equations approach to modelling macroeconomic relationships

- After we have now decided on values for our exogenous variables, we will solve the model in order to provide us with forecasts of the future.
  - Of course we no longer have the choice of a **static forecasting** for the future (as we don't know the true values of the endogenous variables for 2014 – 2020).
  - We hope that our exogenous variables are estimated realistically for the period.

After doing this, we have forecasts into the future!

The next step should be obvious…

The system of equations approach to modelling macroeconomic relationships

- We can use the current forecasts – and compare them to different scenarios, within which we tinker with the values that we specified.

- E.g., we can compare our GDP forecasts with, e.g., interest rates set at 5%, with an interest rate that increases every quarter by 50 bps.
- This allows us to assess what impact the increase in interest rates would have on GDP within our dynamic model setup.
- This we call scenarios, and Eviews will give aliases to variables in different scenarios.

The END

The rest of the steps and interpretations will now be explained in the practical class where we will fit these models.

Please consult the detailed tutorial memos for further information relevant to this session – the tut solutions are treated as equally important for exam purposes.