ECON 4160, Spring term 2013. Lecture 10

FIML examples. Model based macroeconomic forecasts

Ragnar Nymoen

Department of Economics

30 Oct 2013

Today:

- Follow up on the use of multi-equation-dynamic modelling and FIML by a couple of examples
- An introduction to the theory of model based forecasting in economics
- Unfortunately, there is very little about forecasting in the books, even though forecasting is one of the main purposes for formulating VARs and models of the VAR
- However, there is a short section on forecast errors on page 103-104 in DK that links up well with the "forecast theorem" part of this lecture.

The NPC system I

- Lecture 8 we presented GMM estimates of the New Keynesian Phillips curve (NPC) equation
- We saw that estimation issues were clarified by a "completing" system, that we can call the NPC system
- We write it here as:

$$\Delta p_t = b_{p1}^f E_t(\Delta p_{t+1}) + b_{p1}^b \Delta p_{t-1} + b_{p2} x_t + \epsilon_{pt} \quad (1)$$

$$\Delta p_{t+1} = E_t(\Delta p_{t+1}) + e_{t+1} \quad (2)$$

$$x_t = b_{x1} x_{t-1} + \epsilon_{xt}, \; -1 < b_{x1} < 1 \quad (3)$$

where e_{t+1} is the (rational) expectation error of the conditional forecast of Δp_t based on period t information. Assume e_{t+1} to be uncorrelated with ϵ_{pt} and ϵ_{xt}

The NPC system II

- In lecture 8, we assumed that $Cov(\epsilon_{pt}, \epsilon_{xt}) = 0$, now we allow $Cov(\epsilon_{pt}, \epsilon_{xt}) \neq 0$.
- In terms of observables, the dynamic multi-equation model is:

$$\Delta p_t = b_{p1}^f \Delta p_{t+1} + b_{p1}^b \Delta p_{t-1} + b_{p2} x_t + \epsilon_{pt} + b_{p1}^f e_{t+1} \quad (4)$$

$$x_t = b_{x1} x_{t-1} + \epsilon_{xt}, \; -1 < b_{x1} < 1 \quad (5)$$

- Since the NPC equation (4) does not exclude any variables from the system, it cannot be identified.

The NPC system III

- Since most estimations of the NPC, like those cited in Lecture 8, use several overidentifying instruments, the relevant competing model is not (4)-(5), but

$$\Delta p_t = b_{p1}^f \Delta p_{t+1} + b_{p1}^b \Delta p_{t-1} + b_{p2} x_t + \epsilon_{pt} + b_{p1}^f e_{t+1} \quad (6)$$

$$x_t = b_{x1} x_{t-1} + \mathbf{b}_{x2} \mathbf{z}_t + \epsilon_{xt}, \quad (7)$$

  where $\mathbf{z}_t$ is vector with instruments: Typically x_{t-2}, up to five lags of Δp_t and often one of more proxies for output-gap.
- In the NPC literature it is not very clear where the instruments "come from", they do not follow from the theory that unerpin (6) for example. An interpretation maybe that (7) is a reduced form derived form the rest of the model (ie., everything else apart from the NPC equation?)

The NPC system IV

- That said, the point here is to show an example of equation with lead-variables can formulated as part of a SEM and estimated by FIML.
- Example in class.

Suicide rates and unemployment I

- There is a literature (in both sociology and economics) about the "endemic" nature of suicide
- Here: The relationship between the male suicidal rate in Norway (variable name $menns_t$) and the unemployment rate U_t.
- We have annual data from 1905.
- $\mathbf{y}_t' = (Lmenns_t, LU_t)$ where the L denoted the natural logarithm
- The VAR that represents the dynamic system (URF) is:

$$\mathbf{y}_t = \sum_{i=1}^{3} \mathbf{\Pi}_i \mathbf{y}_{t-i} + \mathbf{\Gamma}_0 \mathbf{z}_t + \varepsilon_t \tag{8}$$

where $\mathbf{z}_t$ is a vector with individual year-dummies (1921 and 1945 for example). Graph data in class.

Suicide rates and unemployment II

- (8) is a 2-dimensional VARX(3) with $E(\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}$ (not diagonal by assumption)
- We will first show an example of SURE estimation for this case
- Attempt a dynamic model of (8) based on the idea that $LU_t \rightarrow Lmenns_t$, (triangular contemporaneous matrix) and estimate by FIML to discover that structural disturbances in $\boldsymbol{\epsilon}_t^{'} = (\epsilon_{1t}, \epsilon_{2t})$ can be correlated when we interpret the model with triangular contemporaneous matrix as a restricted SEM model (rather than a model made up of conditional model for suicide and a marginal model for LU).

Terminology and essential assumption for forecasting I

- Surprisingly many insights about model based forecasting can be gained by thinking trough the simplest case where a multi-period forecasts (also called dynamic forecast) $\hat{Y}_{T+h}$ $(h = 1, 2, .., H)$ can be based on the gaussian AR(1):

$$Y_t = \phi_0 + \pi_1 Y_{t-1} + \varepsilon_t \tag{9}$$

$$\varepsilon_t \sim N(0, \sigma^2),\ t = 1, 2, \dots, T \tag{10}$$

- The history of Y_t up to and including period T, Y_t $(t = 1, 2, .., T)$, is an important part of the **information set** of the forecast $\hat{Y}_{T+h}$ $(h = 1, 2, .., T + H)$.
- We call H the forecast horizon.

Terminology and essential assumption for forecasting II

- There is a fundamental difference between the statement that (9)-(10) is a model that we assume to hold within the sample, (up to and including period T);
- And another statement, saying that the model holds for the forecast period: $Y_{t+h}(h = 1, 2, .., H)$.
- One reason for drawing this distinction is that the first statement can be evaluated empirically (how?), while there is no known way (apart from a crystal ball; or by asking a princess who speaks with angels) to assess the second statement: Namely that **the model continues to hold also in the forecast period**.
- Nevertheless, that assumption is underlies
  - All model based macroeconomic forecasting

Terminology and essential assumption for forecasting III

- The rational expectations assumption in macroeconomic theory (or that the agents have the crystal ball, the princess ,or both).

Optimal forecast theorem I

- Assume that
  - the gaussian AR(1) model holds for $t = 1, 2, \ldots, T, T+1, \ldots T+H$
  - and that we know the parameters π_0, π_1, σ^2
  - and that we know the history of the time series Y_t $(t = 1, 2, .., T)$ without measurement errors
- The second abstracts from the "estimation problem" and the third from the "real-time problem" (The observed Y_T that we base our forecast on is a preliminary data release for the true Y_T) of practical forecasting
- Nevertheless, only the first is really fundamental.

Optimal forecast theorem II

- Next, another very important assumption: That the forecasting **loss-function** is quadratic so that we are interesting in minimizing the **mean of squared forecast errors** (MSFE):

$$MSFE = \frac{1}{H}\sum_{h=1}^{H}(Y_{T+h} - \hat{Y}_{T+h|T})^2 \qquad (11)$$

- The theorem says that **the optimal forecasts are the conditional expectations based on the period T information set** $(\mathcal{I}_T)$:

$$\hat{Y}_{T+h|T} \equiv E(Y_{T+h} \mid \mathcal{I}_T),\ h = 1, 2, .., H \qquad (12)$$

Proof is not difficult, but dropped here.

Dynamic forecast (dynamic simulation) I

- To find $\hat{Y}_{T+h|T} \equiv E(Y_{T+h} \mid \mathcal{I}_T)$ we need to solve the model with T as the initialization period.

$$Y_{T+h} = \pi_0 \sum_{i=0}^{h-1} \pi_1^i + \pi_1^h Y_T + \sum_{i=0}^{h-1} \pi_1^i \varepsilon_{T+h-i} \qquad (13)$$
$$h = 1, 2, .., H$$

- And then invoke the gaussian white-noise assumption:

$$\hat{Y}_{T+h|T} = \pi_0 \sum_{i=0}^{h-1} \pi_1^i + \pi_1^h Y_T \qquad (14)$$
$$h = 1, 2, .., H$$

Dynamic forecast (dynamic simulation) II

- Of course, we can obtain the same sequence of forecasts by **dynamic simulation**, which amounts to

$$\begin{aligned}
\hat{Y}_{T+1|T} &\equiv E(Y_{T+1} \mid \mathcal{I}_T) \\
&= \pi_0 + \pi_1 Y_T \\
\hat{Y}_{T+2|T} &= \pi_0 + \pi_1 \hat{Y}_{T+1|T} \\
&= \pi_0(1 + \pi_1) + \pi_1^2 Y_T \\
\hat{Y}_{T+3|T} &= \pi_0 + \pi_1 \hat{Y}_{T+2|T} \\
&= \pi_0(1 + \pi_1 + \pi_1^2) + \pi_1^3 Y_T
\end{aligned}$$

and so on

Dynamic forecasts equilibrium-correct! I

- The above formulae hold even if for example $\pi_1 = 1$, so that Y_t is not stationary. A case that we will come back to in Lecture 11.
- However for the stationary case $-1 < \pi_1 < 1$ we have the interesting result that

$$\hat{Y}_{T+H|T} \underset{H \to \infty}{\rightarrow} E(Y_t) \tag{15}$$

saying that the dynamic forecasts in (14) converge asymptotically to the unconditional expectation of Y_t. Since $E(Y_t)$ represents the equilibrium value of Y_t, define

$$Y^* \equiv E(Y_t) = \frac{\pi_0}{1 - \pi_1}, \text{ iff } -1 < \pi_1 < 1 \tag{16}$$

Dynamic forecasts equilibrium-correct! II

- In fact, an inspired re-expression of (14) allows us to write the dynamic forecasts as

$$\hat{Y}_{T+h|T} = Y^* + \pi_1^h(Y_T - Y^*) \text{ iff } -1 < \pi_1 < 1 \qquad (17)$$
$$h = 1, 2, .., H$$

  showing that a forecast from a stationary AR equilibrium-corrects.
- This point, in particular, generalizes to VARs and therefore to forecasts from econometric models of the VAR: Model based forecasts always show a tendency of equilibrium correction. Strength depend on the multivariate counterpart to π_1^h.
- (17) also tells us when a model based forecast is most usefull (informative).

Forecast errors I

- 1-step ahead forecast error:

$$Y_{T+1} - \hat{Y}_{T+1|T} = \varepsilon_{T+1} \tag{18}$$

- $h-$step ahead forecast errors:

$$Y_{T+h} - \hat{Y}_{T+h|T} = \sum_{i=0}^{h-1} \pi_1^i \varepsilon_{T+h-i} \tag{19}$$
$$h = 1, 2, .., H$$

Forecast errors II

- Expectation and variances:

$$E(Y_{T+h} - \hat{Y}_{T+h|T} \mid \mathcal{I}_T) = 0 \text{ for all } h \tag{20}$$

and that

$$Var(Y_{T+H} - \hat{Y}_{T+H}) \underset{H\to\infty}{\to} Var(Y_t) \tag{21}$$

This also generalizes to stationary VARs and models of VARs, meaning that:

- Forecasts are never biased (systematically over or under)
- The prediction intervals typically increase as functions of the forecast horizon, but stabilizes as the forecast horizon gets longer.

Summary of our forecasting theory I

- Forecasts, like rational expectation, should always be right on average
- The prediction intervals show the correct degree of uncertainty of forecasting Y_t
- As a result of this, **forecast failures** (realization that are outside the prediction intervals) should be rare in economic forecasting
- Are they?

Main problem: Structural breaks in the forecast period I

- The forecasting theory was based on assumptions that are not correct in practice:
  1. We know the structure of the true model (it is a AR(1) in our example)
  2. We know the parameter values
  3. There are no real-time data problems
  4. There are no structural breaks in the forecast period
- 1. and 2 can be resolved by econometrics and methodology!
- 3. *is* a nuisance but mostly for short forecasting horizons (why?).
- The main problem is 4. Forecasts are often damaged by structural breaks in the real world economy that we attempt to forecast by models that don't contain those breaks!

Forecasting bias due to structural breaks I

- Look again at the AR(1) example
- Assume that there is a structural break: $E(Y_t)$ changes from Y^* to $(Y^* + d)$ $(d > 0)$ in period $T + 1$. The true conditional expectations will then follow

$$E(Y_{T+h} \mid Y_T) = (Y^* + d) + \pi_1^h(Y_T - (Y^* + d)) \quad (22)$$

but our model based forecast will still be

$$\hat{Y}_{T+h|T} = Y^* + \pi_1^h(Y_T - Y^*) \text{ iff } -1 < \pi_1 < 1 \quad (23)$$

and the forecast error will have expectation:

$$E(Y_{T+h} - Y_{T+h} \mid \mathcal{I}_T) = (1 - \pi_1^h)d \neq 0 \text{ for all } h$$

so that there is an systematic **bias**

Forecasting bias due to structural breaks II

- The problem is that the forecast equilibrium correct to the *wrong* equilibrium
- Example: $d = 1.5\%$ and $\pi_1 = 0.5$
  - bias_{T+1}: 0.75%
  - bias_{T+3}: 1.12%
  - bias_{T+3}: 1.31%
- Of course, after one period has passed, we can produce a new forecast that conditions on Y_{t+1}, the period where the shock occurs
- Even if the break is now "in" the initialization period, the forecast is biased towards Y^* as before:

$$\hat{Y}_{T+1+h|T+1} = Y^* + \pi_1^h(Y_{T+1} - Y^*)$$

Forecasting bias due to structural breaks III

- Without any intervention: Model based forecasts are not good at adapting to structural breaks.
- It is not before we correct Y^* to $(Y^* + d)$, that the biases will be removed.

Parameters of interest when forecasting I

- When we estimate a model for policy purposes or for testing a hypothesis, the parameters of interest are the **derivative coefficients**: regression coefficients or partial derivatives of a structural equation
- When the purpose is forecasting, the parameters of interest are the **conditional means** of the endogenous variables. These parameters can display breaks even though (like in our example) the derivative coefficients do no break—it is the break in the *Constant* that often damage forecasts.
- On the positive side, this means that a model can be "good for policy analysis", even though it has forecasted badly.

Keynesian type macro model

- Medium term macro model (cf simultaneous equation bias)
  - C_t : private consumption in year t (in constant prices, eg.2010)
  - GDP_t, TAX_t and I_t are *gross domestic product*, *net taxes* and *investments and gov.exp.*
  - a - e are parameters of the macroeconomic model
  - ϵ_{Ct} and ϵ_{TAXt} are independent disturbances with *classical properties* conditional on I_t and C_{t-1}.

$$C_t = a + b(GDP_t - TAX_t) + cC_{t-1} + \epsilon_{Ct} \quad (24)$$

$$TAX_t = d + eGDP_t + \epsilon_{TAXt} \quad (25)$$

$$GDP_t = C_t + I_t \quad (26)$$

- C_t, GDP_t and TAX_t are endogenous, C_{t-1} is predetermined.
- Assume that I_t is strictly exogenous with $E(I_t) = \mu_I$ and $Var(I_t) = \sigma_I^2$. For simplicity, we will use

$$I_t = \mu_I + \epsilon_{It} \quad (27)$$

where is independent of ϵ_{Ct} and ϵ_{TAXt}, and has classical properties conditional on I_t and C_{t-1}.

Reduced form

- Suppose that our purpose is to forecast C_{T+1}, $C_{T+2}, \ldots, C_{T+H}$ based on data up to and including period T.
- This implies that we generate forecasts from the reduced form equation of C_t

$$C_t = \underbrace{\frac{a+bd}{(1-b(1-e))}}_{\beta_0} + \underbrace{\frac{b(1-e)}{(1-b(1-e))}}_{\beta_1} I_t + \underbrace{\frac{c}{(1-b(1-e))}}_{\beta_2} C_{t-1} + \underbrace{\frac{\epsilon_{Ct} - (be)\epsilon_{TAXt}}{(1-b(1-e))}}_{\varepsilon_t}$$

- This is an ARDL model and β_j ($j = 0, 1, 2$) are reduced form coefficients that can be estimated by OLS from a sample $t = 1, 2, \ldots, T$.
- Alternatively, because we are also interested in the structural parameters of the model, we estimate the structural equations with 2SLS or FIML, and obtain the (restricted) reduced form parameters.

1-step ahead forecast I

- Since

$$E(\varepsilon_{T+1} \mid C_T,\ I_T) = 0$$

the model consistent best forecast for $T+1$ becomes:

$$E(C_{T+1} \mid C_T, I_T) = \beta_0 + \beta_1 E(I_{T+1} \mid C_T, I_T) + \beta_2 C_T \quad (28)$$

Next, from the exogeneity of I_t:

$$E(I_{T+1} \mid C_T, I_T) = E(I_{T+1}) = \mu_I$$

so the forecast for C_{T+1} can be written as

$$E(C_{T+1} \mid C_T, I_T) = \beta_0 + \beta_1 \mu_I + \beta_2 C_T$$

1-step ahead forecast II

- Note that I_{T+1} has been forecasted by $E(I_t) = \mu_I$
  - Reminds us that it necessary to also forecast the exogenous variable!
  - In practice forecasters often use subjective forecasts for exogenous variables, and often present alternatives "scenarios".
  - Here we "keep it clean" and have used the mathematical expectation consistent with the model assumptions.
- The practical problem with using $E(C_{T+1} \mid C_T, I_T)$ as a forecast is that the parameters are unknown.

1-step ahead forecast III

- In practice we therefore replace $E(C_{T+1} \mid C_T, I_T)$ by the estimated expectation:

$$\hat{C}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_I + \hat{\beta}_2 C_T \tag{29}$$

where $\hat{\beta}_0$,$\hat{\beta}_1$,$\hat{\beta}_2$ are OLS estimates or derived from FIML estimates of the full structural model.

h-period ahead dynamic forecasts

Define (to simplify notation)

$$\hat{\gamma} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_I$$

Generally, for *forecast horizon h*:

$$\hat{C}_{T+h} = \sum_{j=0}^{h-1} \hat{\beta}_2^j \hat{\gamma} + \hat{\beta}_2^h C_T \qquad h = 1, 2, \ldots, H \tag{30}$$

Long-horizon forecast I

If $-1 < \hat{\beta}_2^h < 1$ we get from

$$\hat{C}_{T+h} = \sum_{j=1}^{h-1} \hat{\beta}_2^j \hat{\gamma}_T + \hat{\beta}_2^h C_T \qquad h = 1, 2, \ldots, H$$

that

$$\hat{C}_{T+h} \underset{h \longrightarrow \infty}{\longrightarrow} \widehat{E(C_t)} = \frac{\hat{\gamma}}{1 - \hat{\beta}_2}$$

exactly as in the first example.

Forecast errors: Bias

$$C_{t+h} - \hat{C}_{T+h} = (\gamma - \hat{\gamma}) \sum_{j=0}^{h-1} \hat{\beta}_2^h + (\beta_2^h - \hat{\beta}_2^h) C_T + \sum_{j=0}^{h-1} \beta_2^j \varepsilon_{t+j}$$

Bias:

$$E\left(C_{t+h} - \hat{C}_{T+h}\right) = \sum_{j=0}^{h-1} E\left[(\gamma - \hat{\gamma})\hat{\beta}_2^h\right] + C_T E(\beta_2^h - \hat{\beta}_2^h) \;\; h = 1, 2, \ldots H \tag{31}$$

- Cannot prove that $E\left(C_{t+h} - \hat{C}_{T+h}\right) = 0$ for any horizon h.
- But biases can be small if the estimation sample period is sufficiently large (cf. Lecture 15)

Forecast errors: Variance

$$C_{t+h} - \hat{C}_{T+h} = (\gamma - \hat{\gamma}) \sum_{j=0}^{h-1} \hat{\beta}_2^h + (\beta_2^h - \hat{\beta}_2^h) C_T + \sum_{j=0}^{h-1} \beta_2^j \varepsilon_{t+j}$$

Bias:

$$Var\left(C_{t+h} - \hat{C}_{T+h}\right) = Var\left[(\gamma - \hat{\gamma}) \sum_{j=0}^{h-1} \hat{\beta}_2^h\right] + \sigma^2 \sum_{j=0}^{h-1} \beta_2^{2j} \quad h = 1, 2, \ldots H \tag{32}$$

- The first part corresponds to the "estimation uncertainty"
- That part will be small if the sample period is sufficiently large
- This means that the second part of (32) will dominate the variance of the forecast error, and:

$$Var\left(C_{t+h} - \hat{C}_{T+h}\right) \underset{h \longrightarrow \infty}{\longrightarrow} Var(C_t) = \frac{\sigma^2}{1 - \beta_2^2}$$

- We expect that the variance of the forecast error converge to the theoretical variance of the forecasted variable.

Illustrating the forecasting theory I

Assume that the parameters of the macro model are as in:

$$C_t = 0 + 0.5(GDP_t - TAX_t) + 0.60C_{t-1} + \epsilon_{Ct} \quad (33)$$

$$TAX_t = -20 + 0.5GDP_t + \epsilon_{TAXt} \quad (34)$$

$$I_t = 100 + \epsilon_{It} \quad (35)$$

$$GDP_t = C_t + I_t + A_t - B_t \quad (36)$$

but that we use the estimated reduced form ARDL for C_t to forecast C_{t+j}.

- Generate data for 1-111.
- Use $t = 1, 2, \ldots, 101$ to estimate structural model using CFIML

Illustrating the forecasting theory II

- Forecast C_{101+h} , $h = 1, 2, \ldots, 10$ from after estimation of stuctural equations by CFIML
- Note that have added A_t (export) and B_t (Import) for more realism in the notation. But we regard them as deterministic variables, that are perfectly predictable, so they do not contribute to the forecast errors.

Three stimation and forecasting exercises:

- No breaks. Forecast from period 102
- $E(I_t)$ reduced from 100 to 95 permanently in period 102. Forecast from period 102.
- Estimate to period 104 and forecast from period 105.