# EC408 Topics in Applied Econometrics

B Fingleton, Dept of Economics, Strathclyde University

# Prediction

- One of the main advantages of regression, VAR and VECM analysis is its ability to PREDICT or FORECAST
- Given a model, we can use the model to estimate the value of the lhs variable that would occur if the rhs variables take specific values
- Typically we would estimate a model and then use that model to predict the lhs variable for some points in time in the future
- However this assume that the relationship between lhs and rhs variables remains the same over the forecast period
- And that our model is correctly specified

# the vector error correction representation

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \ldots + \phi_p \Delta Y_{t-p} + P_0 Y_{t-1} + U_t \qquad (72)$$

Where $\Delta Y_t = Y_t - Y_{t\text{-}1}$ is a g by 1 vector of differences at time t of g endogenous variables. The terms $\Delta Y_{t\text{-}1} = Y_{t\text{-}1} - Y_{t\text{-}2}, \Delta Y_{t\text{-}2} = Y_{t\text{-}2} - Y_{t\text{-}3}, \ldots, \Delta Y_{t\text{-}p} = Y_{t\text{-}p} - Y_{t\text{-}p-1}$ are the lagged differences. There are p lags, so $\phi_j$ applies to the j'th lag. It is a g x g matrix of coefficients to be estimated. Also $U_t$ is a g x 1 vector of error terms. $P_0$ is a g x g matrix which is referred to as the (restricted) long-run matrix.

# the vector error correction representation

let the number of lags p = 1, then equation (72) becomes

$$\Delta \mathrm{Y}_t = \phi_1 \Delta \mathrm{Y}_{t-1} + P_0 \mathrm{Y}_{t-1} + \mathrm{U}_t$$

for three variables, say $Y_{1t}, Y_{2t}, Y_{3t}$

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} \Delta Y_{1t-1} \\ \Delta Y_{2t-1} \\ \Delta Y_{3t-1} \end{bmatrix} + \begin{bmatrix} P_{011} & P_{012} & P_{013} \\ P_{021} & P_{022} & P_{023} \\ P_{031} & P_{032} & P_{033} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ Y_{3t-1} \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \\ U_{3t} \end{bmatrix}$$

so that for variable $Y_1$

$$\Delta Y_{1t} = \phi_{11} \Delta Y_{1t-1} + \phi_{12} \Delta Y_{2t-1} + \phi_{13} \Delta Y_{3t-1} + P_{011} Y_{1t-1} + P_{012} Y_{2t-1} + P_{013} Y_{3t-1} + U_{1t}$$

and so on for $Y_2, Y_3$

# the vector error correction representation

A mathematically equivalent way to write out this model is in terms of levels and lagged levels,

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{bmatrix} = \begin{bmatrix} 1+P_{011} & P_{012} & P_{013} \\ P_{021} & 1+P_{022} & P_{023} \\ P_{031} & P_{032} & 1+P_{033} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ Y_{3t-1} \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \\ U_{3t} \end{bmatrix} = \begin{bmatrix} \pi_{111} & \pi_{112} & \pi_{113} \\ \pi_{121} & \pi_{122} & \pi_{123} \\ \pi_{131} & \pi_{132} & \pi_{133} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ Y_{3t-1} \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \\ U_{3t} \end{bmatrix}$$

so that for variable $Y_1$

$$Y_{1t} = (1+P_{011})Y_{1t-1} + P_{012}Y_{2t-1} + P_{013}Y_{3t-1} + U_{1t} = \pi_{111}Y_{1t-1} + \pi_{112}Y_{2t-1} + \pi_{113}Y_{3t-1} + U_{1t}$$

$$Y_{2t} = P_{021}Y_{1t-1} + (1+P_{022})Y_{2t-1} + P_{023}Y_{3t-1} + U_{2t} = \pi_{121}Y_{1t-1} + \pi_{122}Y_{2t-1} + \pi_{123}Y_{3t-1} + U_{2t}$$

$$Y_{3t} = P_{031}Y_{1t-1} + P_{032}Y_{2t-1} + (1+P_{033})Y_{3t-1} + U_{3t} = \pi_{131}Y_{1t-1} + \pi_{132}Y_{2t-1} + \pi_{133}Y_{3t-1} + U_{3t}$$

and $\pi_1$ is a g x g matrix of coefficients specific to lag 1

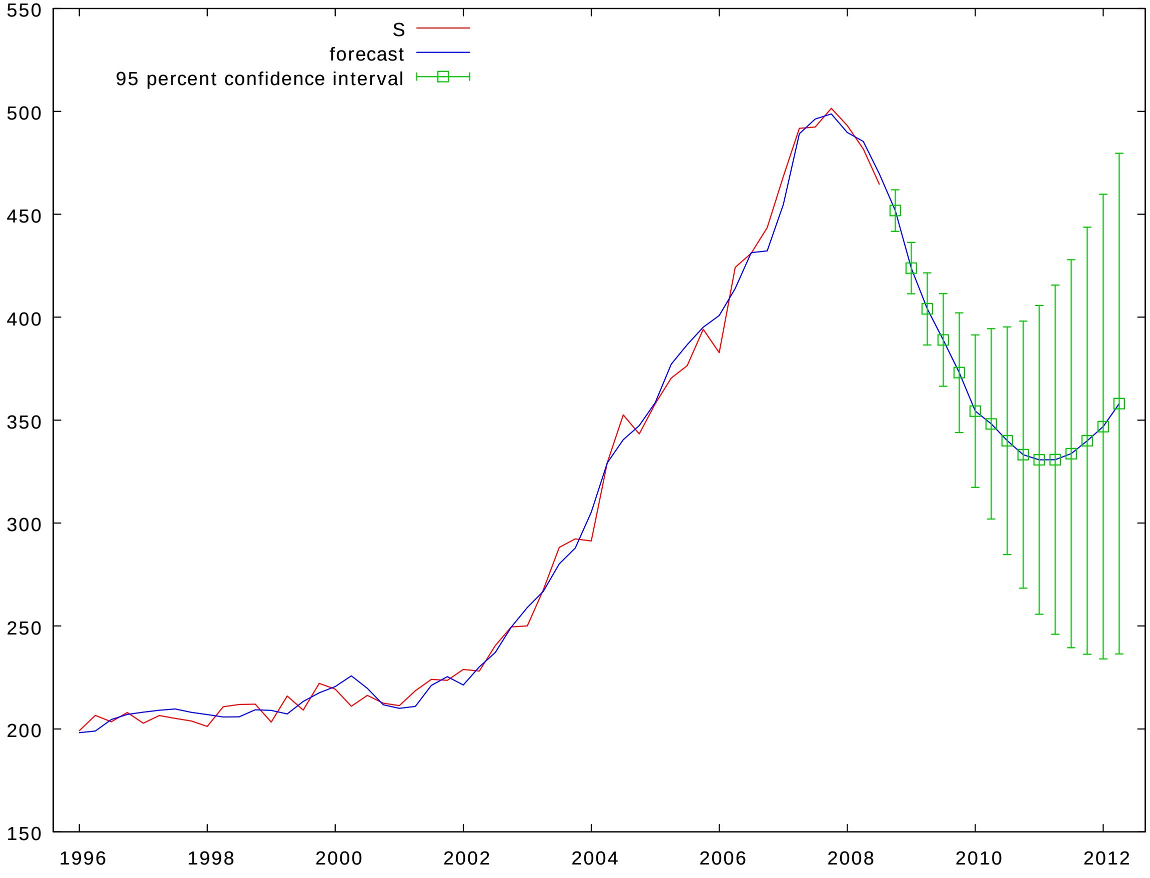
# Johansen's procedure

Given that we have established the rank of $P_0$ and hence the number of cointegrating vectors, we can then move forward in the knowledge that we have a balanced model with stationary variables.

we can obtain estimates of the dependencies within the data that are not spurious, and

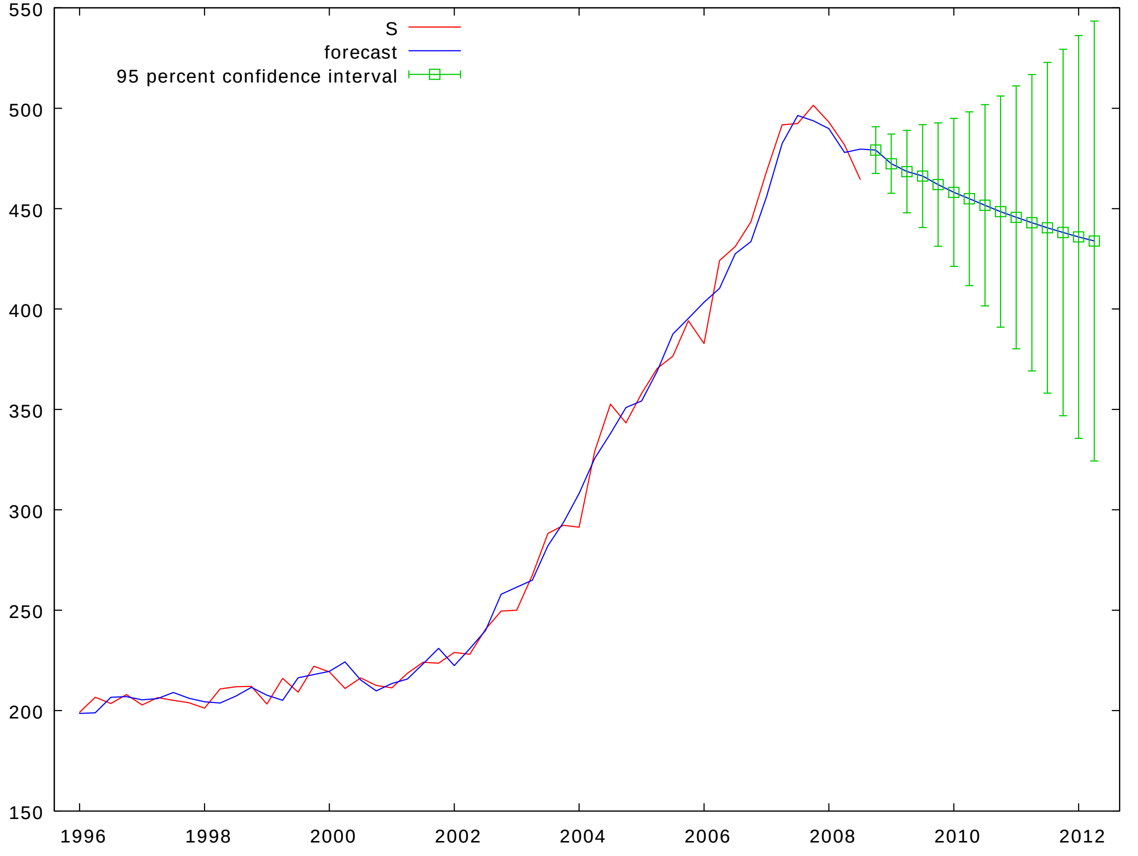
we will ultimately be able to produce more credible forecasts and a richer and more informative picture of the interrelationships between the variables.

# Scotland house prices : dynamic out of sample forecasts



```
VECM system, lag order 5
Maximum likelihood estimates, observations 1984:2-2008:3 (T
= 98)
Cointegration rank = 2
Case 3: Unrestricted constant
Log-likelihood = -1243.4837
```

| | | | | | |
|---|---|---|---|---|---|
| 2006:1 | 382.8 | 400.8 | | | |
| 2006:2 | 424.2 | 414.0 | | | |
| 2006:3 | 431.1 | 431.4 | | | |
| 2006:4 | 443.5 | 432.2 | | | |
| 2007:1 | 468.4 | 454.8 | | | |
| 2007:2 | 491.8 | 489.1 | | | |
| 2007:3 | 492.4 | 496.2 | | | |
| 2007:4 | 501.5 | 498.8 | | | |
| 2008:1 | 493.1 | 489.8 | | | |
| 2008:2 | 481.7 | 485.4 | | | |
| 2008:3 | 464.6 | 469.7 | | | |
| 2008:4 | | 451.8 | 5.16 | 441.7 - | 461.9 |
| 2009:1 | | 423.9 | 6.35 | 411.4 - | 436.3 |
| 2009:2 | | 404.1 | 8.93 | 386.6 - | 421.6 |
| 2009:3 | | 389.0 | 11.47 | 366.5 - | 411.4 |
| 2009:4 | | 373.1 | 14.84 | 344.0 - | 402.2 |
| 2010:1 | | 354.4 | 18.91 | 317.3 - | 391.5 |
| 2010:2 | | 348.2 | 23.59 | 301.9 - | 394.4 |
| 2010:3 | | 340.0 | 28.21 | 284.7 - | 395.3 |
| 2010:4 | | 333.2 | 33.08 | 268.4 - | 398.1 |
| 2011:1 | | 330.7 | 38.26 | 255.7 - | 405.7 |
| 2011:2 | | 330.8 | 43.26 | 246.0 - | 415.6 |
| 2011:3 | | 333.7 | 48.09 | 239.5 - | 428.0 |
| 2011:4 | | 340.0 | 52.93 | 236.2 - | 443.7 |
| 2012:1 | | 346.9 | 57.58 | 234.0 - | 459.7 |
| 2012:2 | | 358.0 | 62.03 | 236.4 - | 479.6 |



550
500
450
400
350
300
250
200
150
1996
1998
2000
2002
2004
2006
2008
2010
2012
S
forecast
95 percent confidence interval

```
VECM system, lag order 3
Maximum likelihood estimates, observations 1983:4-2008:3 (T
= 100)
Cointegration rank = 1
Case 1: No constant
Log-likelihood = -1335.8867
```

| | | | | | |
|---|---|---|---|---|---|
| 2006:1 | 382.8 | 403.3 | | | |
| 2006:2 | 424.2 | 410.3 | | | |
| 2006:3 | 431.1 | 427.5 | | | |
| 2006:4 | 443.5 | 433.6 | | | |
| 2007:1 | 468.4 | 455.8 | | | |
| 2007:2 | 491.8 | 482.5 | | | |
| 2007:3 | 492.4 | 496.4 | | | |
| 2007:4 | 501.5 | 493.8 | | | |
| 2008:1 | 493.1 | 489.9 | | | |
| 2008:2 | 481.7 | 478.0 | | | |
| 2008:3 | 464.6 | 479.6 | | | |
| 2008:4 | | 479.2 | 5.95 | 467.5 - | 490.8 |
| 2009:1 | | 472.4 | 7.52 | 457.6 - | 487.1 |
| 2009:2 | | 468.5 | 10.47 | 448.0 - | 489.0 |
| 2009:3 | | 466.2 | 13.07 | 440.6 - | 491.9 |
| 2009:4 | | 462.0 | 15.68 | 431.3 - | 492.7 |
| 2010:1 | | 458.1 | 18.79 | 421.3 - | 494.9 |
| 2010:2 | | 455.0 | 22.08 | 411.7 - | 498.3 |
| 2010:3 | | 451.7 | 25.58 | 401.6 - | 501.8 |
| 2010:4 | | 448.5 | 29.37 | 390.9 - | 506.1 |
| 2011:1 | | 445.7 | 33.41 | 380.2 - | 511.2 |
| 2011:2 | | 443.0 | 37.64 | 369.2 - | 516.8 |
| 2011:3 | | 440.5 | 42.03 | 358.1 - | 522.9 |
| 2011:4 | | 438.1 | 46.56 | 346.8 - | 529.4 |
| 2012:1 | | 435.9 | 51.19 | 335.6 - | 536.3 |
| 2012:2 | | 433.9 | 55.90 | 324.3 - | 543.4 |

# Starting simply!

Typically, with one independent variable we estimate a regression

$$\hat{Y}_{it} = \hat{b}_0 + \hat{b}_1 X_{it}$$

and use this estimated model to predict what $Y$ will be for future $X$ values

$$\hat{Y}_{it+1} = \hat{b}_0 + \hat{b}_1 X_{it+1}$$

and so on, forecasting to $p$ periods ahead

$$\hat{Y}_{it+p} = \hat{b}_0 + \hat{b}_1 X_{it+p}$$

# Prediction

- however, we should also take account of the inherent uncertainty in this prediction
    - as a result of the fact that $b_0$ and $b_1$ are estimates, not the true values

# Prediction

- assume that we fit the regression
  - $Y_t = b_0 + b_1 time_t + e_t \quad t = 1,\ldots,T$
- with estimates of $b_0$ and $b_1$ it is possible to

predict $Y_{T+1}$ given $time_{T+1}$

$$\hat{Y}_{T+1} = \hat{b}_0 + \hat{b}_1 \, time_{T+1}$$

$$\hat{Y}_{T+2} = \hat{b}_0 + \hat{b}_1 \, time_{T+2}$$

The $\hat{Y}$ ‘s are estimates of E($Y$), the mean of $Y$ at $T$+1, $T$+2 etc

The MSFE is given by the variance of the forecast error

$$E(Y_{T+p} - X_{T+p}\hat{\beta})^2 = E(X_{T+p}\beta + u_{T+p} - X_{T+p}\hat{\beta})^2$$

where $X_{T+p}$ is the observed value of regressors $X$ at time $T+p$

$$= E(u_{T+p}^2) + E(X_{T+p}\beta - X_{T+p}\hat{\beta})^2$$

$$= \sigma^2 + \mathrm{var}(X_{T+p}\hat{\beta})$$

$$= \sigma^2 + X_{T+p}\,\mathrm{var}(\hat{\beta})X'_{T+p}$$

# MSFE

- Variance of forecast error (MSFE) is sum of 2 terms
- Variance of the error term
- Second term increases variance due to the fact that the coefficients are estimates rather than the true values
- This is the penalty paid for our ignorance of the true values of the coefficients

The 95% confidence interval for E($Y$)

Assume the forecast errors are **normally** distributed, then

$$c.i. = 95\% \text{ confidence interval}$$

$$c.i. = \hat{Y}_{T+1} \pm 1.96\, s.e.(Y_{T+1} - \hat{Y}_{T+1})$$

$$c.i. \approx \hat{Y}_{T+1} \pm 1.96 \text{RMSFE}$$

# Summary

- Take special care predicting $Y$ beyond the range of the variables
- always accompany any prediction of $E(Y)$ with its (say 95%) confidence interval

# Prediction

- Notice that we are making several big assumptions
- The regression parameters, and the error variance, estimated for the period 1….*T* remain constant over the period *T*+1….*T*+*p*
- The functional form remains the same
- We 'know' the values of the rhs variables over the period *T*+1….*T*+*p*
- If we do, then we are making what is known as an ex post prediction
- If we do not, then we are making an ex ante prediction. This requires us to first predict the rhs in order to predict the lhs variable
- Needless to say, ex post prediction is much safer than ex ante prediction

# Pseudo out-of-sample forecasts

  - Stock and Watson p. 571
- True out-of-sample forecasting is when forecast without knowing true values of what you are forecasting
- Pseudo is when you hold back some of the data and forecast its values

# Pseudo out-of-sample forecasts

- Allows accuracy of ‘forecasts’ to be checked against real data
- Can be used to measure RMSFE and hence the c.i.
  - Using the standard deviation of the forecast errors in the pseudo forecast
- Allows model comparison, which models are forecasting more accurately?

# Static vs dynamic forecasts

- Static forecasts are one step ahead, based on realized values from the previous period
- dynamic forecasts employ the chain rule of forecasting.
  - For example, if a forecast for Y in 2008 requires as input of Y for 2007, a static forecast is impossible without actual data for 2007. A dynamic forecast for 2008 is possible if a prior forecast can be substituted for Y in 2007.

# dynamic forecasts

- Stock and Watson p. 645 discuss iterated multivariate VAR forecasts – several steps
- Step 1 compute the one period ahead forecasts of all the variables in the VAR
- Step 2 Use those forecasts to compute the 2-period ahead forecasts
- Step 3 use the step 2 forecasts to compute the 3 period ahead forecasts
- And so on, continuing this process iteratively to the forecast horizon

# Dynamic forecast for VAR

The dynamic $h$-step-ahead forecast begins by using the estimated coefficients, the lagged values of the endogenous variables, and any exogenous variables to predict one step ahead for each endogenous variable. Then the one-step-ahead forecast produces two-step-ahead forecasts for each endogenous variable. The process continues for $h$ periods. Because each step uses the predictions of the previous steps, these forecasts are known as dynamic forecasts. See the following sections for information on obtaining forecasts after `svar`:

# Dynamic forecast for VAR

As shown by Lütkepohl (2005, 204–205), the asymptotic estimator of the covariance matrix of the prediction error is given by

$$\widehat{\boldsymbol{\Sigma}}_{\widehat{y}}(h) = \widehat{\boldsymbol{\Sigma}}_y(h) + \frac{1}{T}\widehat{\boldsymbol{\Omega}}(h) \tag{2}$$

where

$$\widehat{\boldsymbol{\Sigma}}_y(h) = \sum_{i=0}^{h-1} \widehat{\boldsymbol{\Phi}}_i \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Phi}}_i'$$

Equation (2) is made up of two terms. $\widehat{\boldsymbol{\Sigma}}_y(h)$ is the estimated mean squared error (MSE) of the forecast. $\widehat{\boldsymbol{\Sigma}}_y(h)$ estimates the error in the forecast arising from the unseen innovations. $T^{-1}\widehat{\boldsymbol{\Omega}}(h)$ estimates the error in the forecast that is due to using estimated coefficients instead of the true coefficients. As the sample size grows, uncertainty with respect to the coefficient estimates decreases, and $T^{-1}\widehat{\boldsymbol{\Omega}}(h)$ goes to zero.

# Dynamic forecast for VAR

If $\mathbf{y}_t$ is normally distributed, the bounds for the asymptotic $(1-\alpha)100\%$ interval around the forecast for the $k$th component of $\mathbf{y}_t$, $h$ periods ahead, are

$$\widehat{\mathbf{y}}_{k,t}(h) \pm z_{(\frac{\alpha}{2})}\widehat{\sigma}_k(h) \tag{3}$$

where $\widehat{\sigma}_k(h)$ is the $k$th diagonal element of $\widehat{\mathbf{\Sigma}}_{\widehat{y}}(h)$.

# Dynamic forecast for vec

Per Lütkepohl (2005, sec. 6.5), `fcast compute` uses

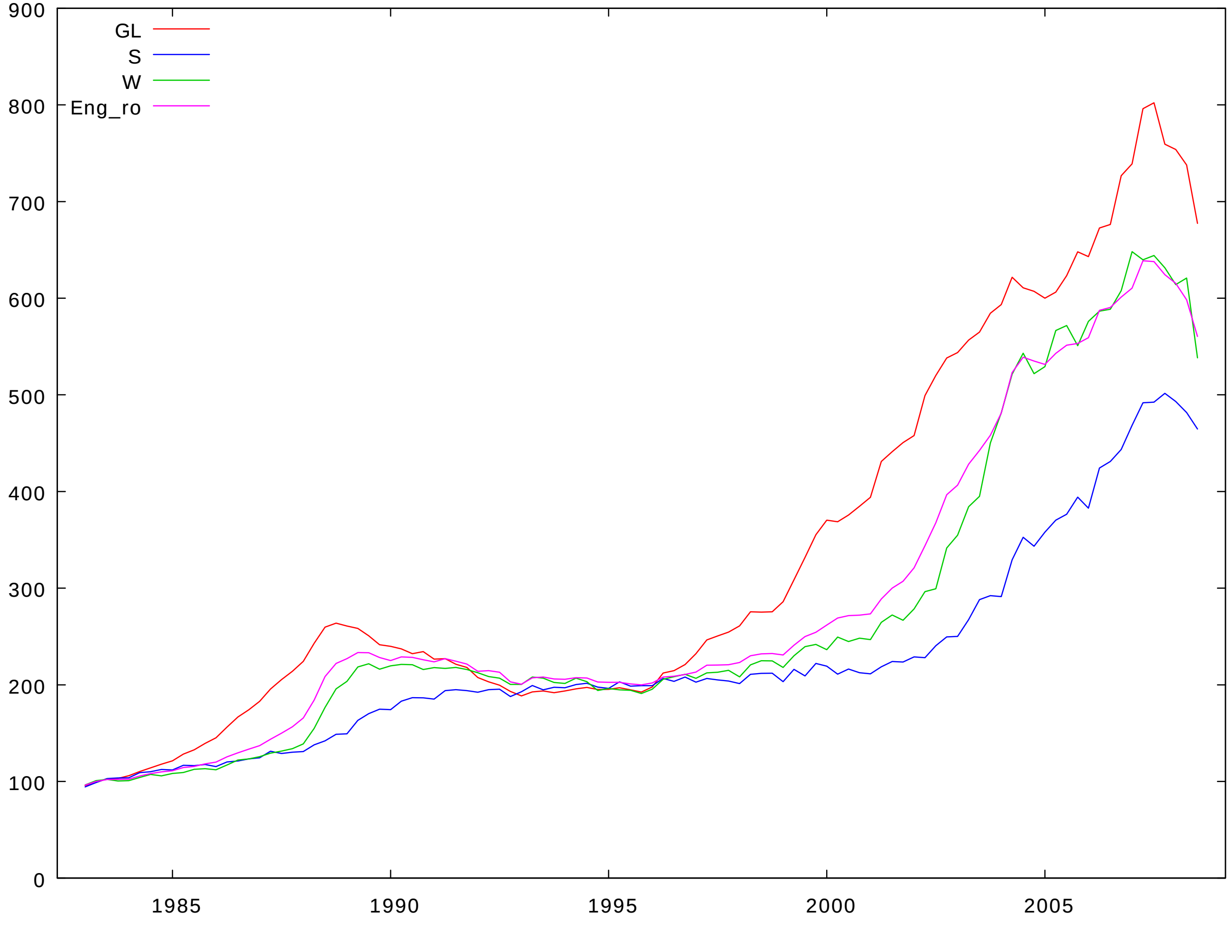
$$\widehat{\boldsymbol{\Sigma}}_{\widehat{y}}(h) = \left(\frac{T}{T-d}\right)\sum_{i=0}^{h-1}\widehat{\boldsymbol{\Phi}}_i\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Phi}}_i$$

where the $\widehat{\boldsymbol{\Phi}}_i$ are the estimated matrices of impulse–response functions, $T$ is the number of observations in the sample, $d$ is the number of degrees of freedom, and $\widehat{\boldsymbol{\Omega}}$ is the estimated cross-equation variance matrix. The formulas for $d$ and $\widehat{\boldsymbol{\Omega}}$ are given in *Methods and formulas* of [TS] **vec**.

The estimated standard errors at step $h$ are the square roots of the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_{\widehat{y}}(h)$.

Per Lütkepohl (2005), the estimated forecast-error variance does not consider parameter uncertainty. As the sample size gets infinitely large, the importance of parameter uncertainty diminishes to zero.

# House price data in GB



# Pseudo out of sample forecasts house price data

VECM system, lag order 5

Maximum likelihood estimates, observations 1984:2-2007:4 (T = 95)

Cointegration rank = 3

Case 1: No constant

# Pseudo out-of-sample dynamic forecasts house price data

For 95% confidence intervals, z(.025) = 1.96

| Obs | Eng_ro | prediction | std. error | 95% confidence interval |
|---|---|---|---|---|
| 2008:1 | 615.425 | 620.347 | 4.5364 | 611.456 - 629.238 |
| 2008:2 | 598.275 | 615.265 | 8.9664 | 597.691 - 632.840 |
| 2008:3 | 560.400 | 622.588 | 13.3821 | 596.359 - 648.817 |

| Obs | GL | prediction | std. error | 95% confidence interval |
|---|---|---|---|---|
| 2008:1 | 753.8 | 770.8 | 9.42 | 752.4 - 789.3 |
| 2008:2 | 737.8 | 807.2 | 14.52 | 778.8 - 835.7 |
| 2008:3 | 677.4 | 843.1 | 19.92 | 804.1 - 882.2 |

| Obs | S | prediction | std. error | 95% confidence interval |
|---|---|---|---|---|
| 2008:1 | 493.1 | 490.8 | 5.10 | 480.8 - 500.8 |
| 2008:2 | 481.7 | 502.8 | 6.12 | 490.8 - 514.8 |
| 2008:3 | 464.6 | 495.5 | 7.76 | 480.3 - 510.8 |

| Obs | W | prediction | std. error | 95% confidence interval |
|---|---|---|---|---|
| 2008:1 | 614.1 | 598.1 | 7.89 | 582.7 - 613.6 |
| 2008:2 | 620.9 | 587.8 | 12.33 | 563.6 - 612.0 |
| 2008:3 | 538.2 | 604.8 | 15.78 | 573.9 - 635.7 |

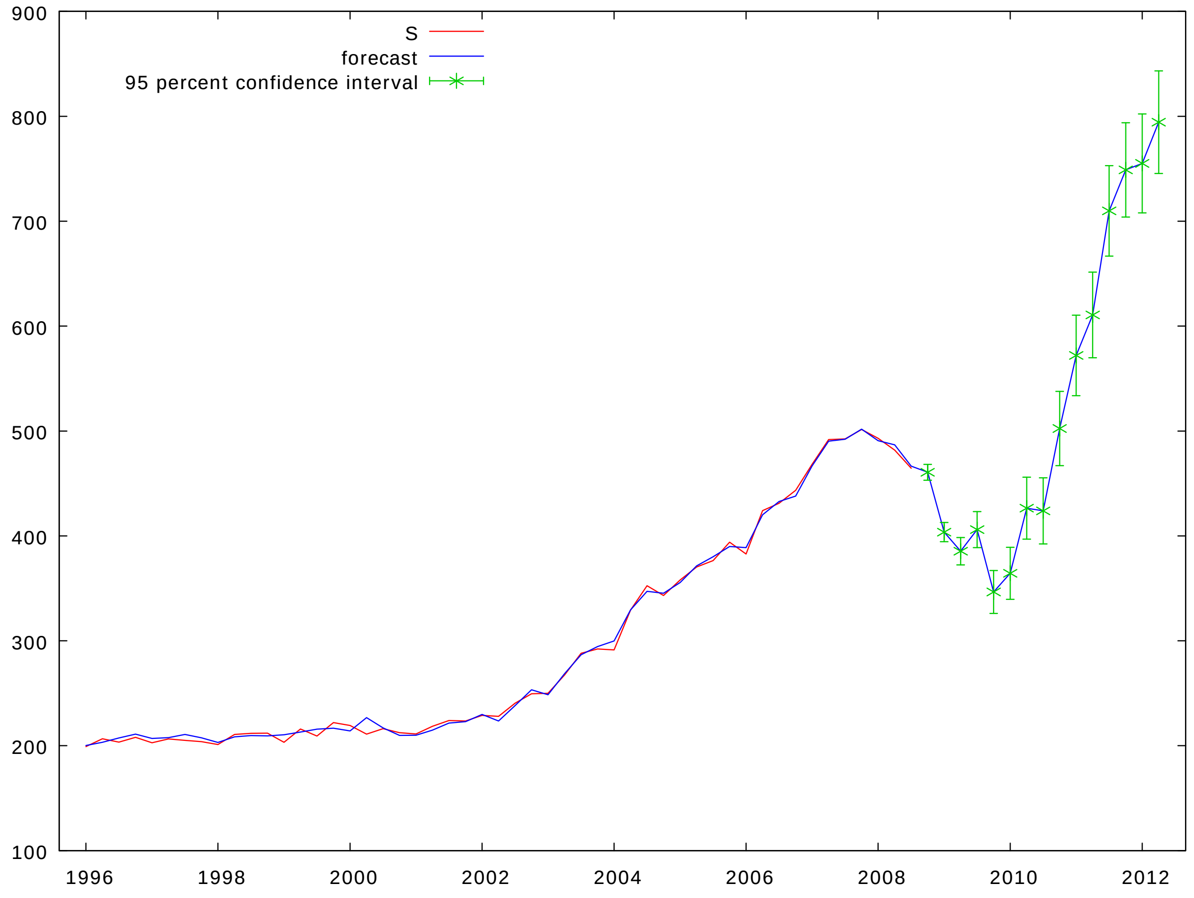
# Which rank?

Johansen test:
Number of equations = 4
Lag order = 10
Estimation period: 1985:3 - 2008:3 (T = 93)

Case 1: No constant

| Rank | Eigenvalue | Trace test | p-value | Lmax test | p-value |
|---|---|---|---|---|---|
| 0 | 0.50571 | 97.402 | [0.0000] | 65.530 | [0.0000] |
| 1 | 0.18398 | 31.872 | [0.0039] | 18.908 | [0.0314] |
| 2 | 0.10767 | 12.964 | [0.0381] | 10.595 | [0.0637] |
| 3 | 0.025151 | 2.3690 | [0.1450] | 2.3690 | [0.1463] |

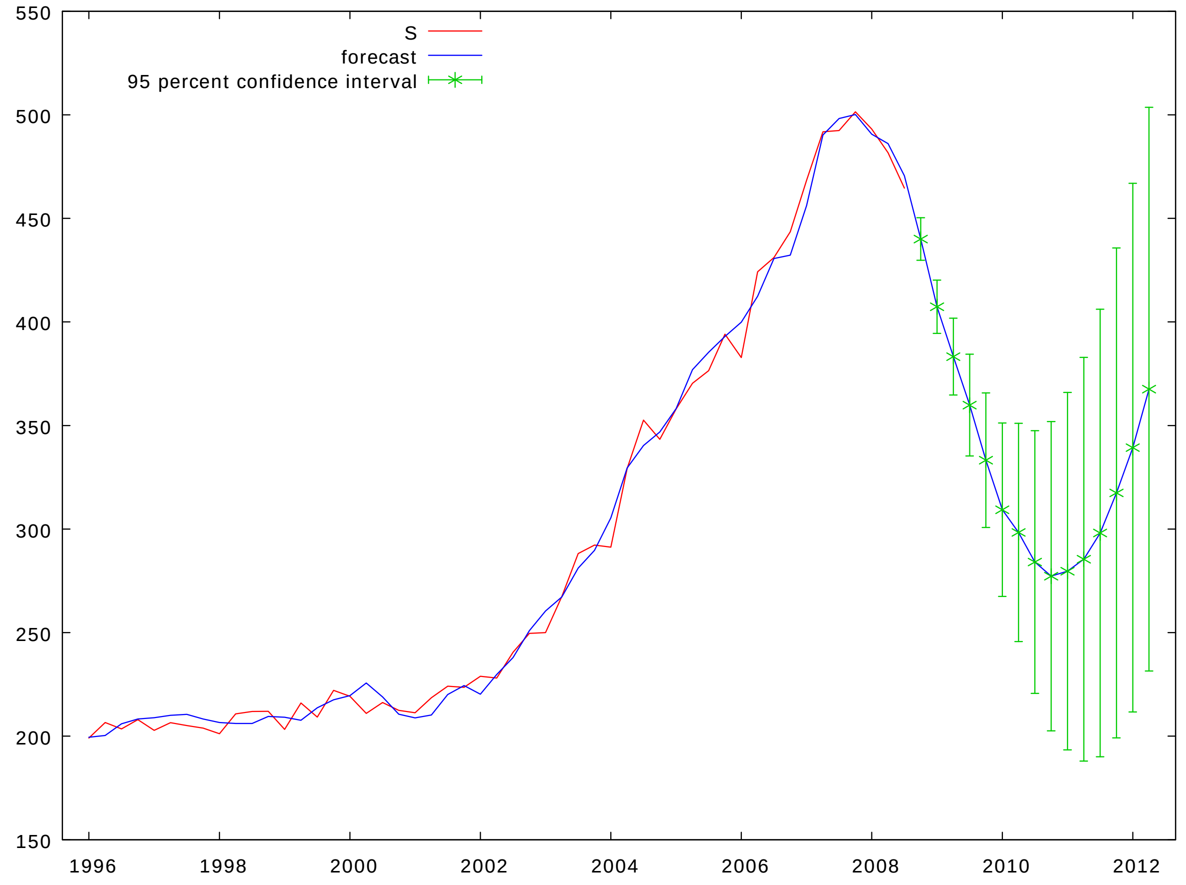
# Scotland’s house prices, VECM, 10 lags, 3 ci vectors



# Scotland's house prices, VECM, 10 lags, 3 ci vectors

| | | | | | |
|---|---|---|---|---|---|
| 2008:3 | 464.6 | 466.5 | | | |
| 2008:4 | | 460.7 | 3.86 | 453.1 - | 468.2 |
| 2009:1 | | 403.6 | 4.68 | 394.4 - | 412.8 |
| 2009:2 | | 385.4 | 6.68 | 372.3 - | 398.5 |
| 2009:3 | | 406.1 | 8.75 | 388.9 - | 423.2 |
| 2009:4 | | 346.6 | 10.41 | 326.2 - | 367.0 |
| 2010:1 | | 364.4 | 12.67 | 339.6 - | 389.2 |
| 2010:2 | | 426.5 | 15.05 | 397.0 - | 456.0 |
| 2010:3 | | 423.9 | 16.08 | 392.4 - | 455.4 |
| 2010:4 | | 502.4 | 18.08 | 467.0 - | 537.9 |
| 2011:1 | | 572.0 | 19.61 | 533.6 - | 610.5 |
| 2011:2 | | 610.6 | 20.82 | 569.8 - | 651.4 |
| 2011:3 | | 709.8 | 22.01 | 666.7 - | 753.0 |
| 2011:4 | | 748.9 | 22.92 | 704.0 - | 793.9 |
| 2012:1 | | 755.1 | 24.07 | 707.9 - | 802.3 |
| 2012:2 | | 794.4 | 24.99 | 745.4 - | 843.4 |

# Scotland's house prices, VECM, 5 lags, 3 ci vectors



# Scotland's house prices, VECM, 5 lags, 3 ci vectors

| | | | | | |
|---|---|---|---|---|---|
| 2008:3 | 464.6 | 470.6 | | | |
| 2008:4 | | 440.0 | 5.25 | 429.7 - | 450.3 |
| 2009:1 | | 407.3 | 6.57 | 394.4 - | 420.2 |
| 2009:2 | | 383.3 | 9.45 | 364.7 - | 401.8 |
| 2009:3 | | 359.8 | 12.53 | 335.3 - | 384.4 |
| 2009:4 | | 333.3 | 16.58 | 300.8 - | 365.8 |
| 2010:1 | | 309.3 | 21.37 | 267.4 - | 351.2 |
| 2010:2 | | 298.4 | 26.88 | 245.7 - | 351.1 |
| 2010:3 | | 284.1 | 32.36 | 220.6 - | 347.5 |
| 2010:4 | | 277.3 | 38.09 | 202.6 - | 351.9 |
| 2011:1 | | 279.7 | 44.03 | 193.4 - | 366.0 |
| 2011:2 | | 285.4 | 49.72 | 188.0 - | 382.9 |
| 2011:3 | | 298.1 | 55.12 | 190.1 - | 406.1 |
| 2011:4 | | 317.5 | 60.33 | 199.2 - | 435.7 |
| 2012:1 | | 339.3 | 65.10 | 211.7 - | 466.9 |
| 2012:2 | | 367.5 | 69.42 | 231.5 - | 503.6 |