DSGE Models: Evaluation and forecasting

Fabio Canova
EUI and CEPR
October 2012
Outline

- Principles of policy evaluation
- In-sample evaluation
- Out-of-sample evaluation
- Evaluation via other features (internal propagation, co-cycles, cointegration).
- Evaluation via VARs.
- Evaluation via loss functions.
References


Del Negro, M. and Schorfheide, F. (2005), ”Policy predictions if the model does not fit”, *Journal of the European Economic Association*.


Favero, C. (2006), Model evaluation in Macroeconomics from Cowles foundation to DSGE models, IGIER manuscript.


1 Principles of policy evaluation: Theory

\[ x = m(p, \beta_m, \eta) \] (1)

\( p \) = policy variable, \( m \) = model, \( \beta_m \) = parameters of model \( m \), \( \eta \) random errors.

- Case 1: Model, parameters, errors are known.

- Set up a loss function \( L(x) \).

- Find the \( p \) which minimizes the loss function.

Unrealistic setup!!
• Case 2: \( \eta \) unknown, but its pdf \( \mu_\eta \) is available.

- \( L(x) \) is a random variable and, for each \( p \), the loss function has a distribution.

- Evaluate \( \mu(L(x)|p, m, \beta_m) \) (the pdf of the loss function) or \( L(\mu((x)|p, m, \beta_m)) \) (the loss of the pdf of \( x \)), e.g. find the \( p \) which gives the \( \mu(L(x)) \) with the lower variability.

- Still unrealistic setup.
Case 3: $\eta, \beta_m$ unknown.

a) Evaluate policies using $\mu(L(x)|p, m, \hat{\beta}_m)$, where the data $d$ is used estimate $\beta_m$.

b) Parameter averaging: use $\mu(L(x)|p, m) = \int \mu(L(x)|p, m, \beta_m) d\mu(\beta_m|d)$ where $d\mu(\beta_m|d)$ is the posterior distribution of $\beta_m$, conditional on the data $d$.

- Parameter uncertainty is small if $T$ is sufficiently large.
• Case 4: $\eta, \beta_m$ and $m$ all unknown. Why is $m$ unknown?

i) Unclear which economic theory one should use.

ii) Different functional forms can represent the same theory.

What can you do in this situation?

a) Use model selection criteria (AIC, BIC, etc.), i.e. use $\mu(L(x)|p, \hat{m}, \hat{\beta}_m)$ where $\hat{\beta}_m$ is chosen after $\hat{m}$ is selected.

Problems: (i) data mining; (ii) pre-testing matters (artificially small s.e.); (iii) a model could be good according to the chosen selection criteria but may have low or zero posterior probability.

b) Do model averaging ($m = 1, \ldots, M$), i.e. use $\mu(L(x)|p) = \sum_m \int \mu(L(x)|p, m, \beta_m)d\mu(\beta_m|m)\mu(m|d)$, where $\mu(m|d)$ is the posterior of model $m$, given the data.
In standard exercises one computes;
\[ E(L(x)|p, m, \hat{\beta}_m) = \int L(x)\mu(x|p, m, \hat{\beta}_m)dx. \]

With model averaging (but not parameter averaging) one computes:
\[ E(L(x)|p, \hat{\beta}_m) = \int L(x)\mu(x|p, \hat{\beta}_m)dx. \]

**Example 1** Suppose we care about the long run effect of a policy choice on the variability of \( x \). One can compute:
\[
\begin{align*}
\text{var}(x_\infty|p, m, \beta_m) \\
\text{var}(x_\infty|p, m) \\
\text{var}(x_\infty|p)
\end{align*}
\]

The latter is the effect of policy on long run variability of \( x \) without assuming that the model selection exercise has identified the correct one.
- Averaging is theoretically OK, but policymakers mainly interested in knowing whether and how alternative assumptions about policy affect loss function. (Calculating expected loss may not be that interesting for them). Alternatives statistics:

1) **Outcome Dispersion**

\[ L(x|m_1, p) - L(x|m_2, p) \]  \hspace{1cm} (3)

2) **Action Dispersion**

\[ L(x|m_1, p(m_1)) - L(x|m_2, p(m_2)) \]  \hspace{1cm} (4)

In 2) can ask: does conditioning the policy on a particular model changes the outcome of the experiment? By how much?

In practice it is common to proceed as in case 3. Policymakers informally do model averaging.
How do we use these principles for estimation purposes?

For estimation purposes, \( p \) is given (e.g. Taylor rule). Interested in measuring model fit. Can use the same ideas:

i) Set up a loss function.

ii) Condition on a model-parameter estimate.

iii) Measure discrepancy.

or

ii’) Have an array of models and/or an array of potential estimates.

iii’) Average over models-estimates and measure discrepancy.
2 DSGE Model evaluation

- Statistical vs. economic evaluation?

  - Cowles: Evaluation = test of overidentifying (statistical) restrictions.

  - Calibration: Evaluation = informal distance of moments with economic interpretation.

  - DSGE-Bayesian: match conditional dynamics, measure credibility of models restrictions. Both statistical and economic evaluation are possible.

  - Central Bank models: what criteria do they use?
2.1 In-sample evaluation

2.2 Graphical evaluation

Let $y_t - \hat{y}_t$ the prediction error of the model. Prediction error should be:

- mean zero, iid (no trend or serial correlation should be detected).

- be homoskedastic (no clear break in the variance should be spotted).

- no shock should have ”unreasonable” variance.
Residuals of a three equation sticky price model. Ok?

- Alternative plot actual and predicted values. Any interesting discrepancies? When?
2.2.1 Statistical tests

- Assume you have available a reference model ("the traditional one") and an alternative one.

- Test whether the Mean square Error (MSE) or Mean Absolute Error (MAE) of two (or more) models is the same.

Let $y^1_t$ and $y^2_t$ be the predicted value of $y_t$ from models $m_1$ and $m_2$. Estimate jointly

\begin{align*}
y_t - y^1_t &= \mu + \epsilon^1_t \quad (5) \\
y_t - y^2_t &= \mu + \epsilon^2_t \quad (6)
\end{align*}

where $\epsilon^1_t, \epsilon^2_t$ have the same variance, $\sigma^2$. Estimate the mean and the variance of each equation separately. Use a $\chi^2(2)$ test to verify if the restrictions hold (if they do than $MSE = \mu^2 + \sigma^2$ is the same for the two models, if they don’t, the MSE is different).
• Compute unbiasedness regressions

\[ y_t = a + by_t^* + u_t \]  \hfill (7)

\( y_t^* \) is the predicted value. Ideally \( a = 0, b = 1 \) for a ”good” model.

• Compute predictive regressions

\[ y_t = ay_t^b + by_t^* + e_t \]  \hfill (8)

\( y_t^* \) the predicted value for the (structural) model, \( y_t^b \) is the predicted value from a baseline (time series) model.

Check whether \( b \neq 0 \), i.e. does the new model adds information to the previous model?
- Estimate an unobservable factor model

\[ y_t^* = \delta + \Lambda f_t^* + \epsilon_{1,t} \]  
\[ y_t = a + b f_t^* + \epsilon_{2,t} \]

(9)  
(10)

\( f_t^* \) is the (unknown) predictable part of \( y_t \). Here \( y_t^* \) is not an unbiased predictor for \( y_t \) but only a noisy measure of its predictable component. Does the first equation adds information to the estimates in the second?

- Add predicted values in a VAR and test significance (similar in spirit to predictive regressions - here looks at lagged info).

\[ y_t = A(\ell)y_{t-1} + B(\ell)y_{t-1}^* + u_t \]

(11)

Test \( B(\ell) = 0 \), jointly or separately for each equation.

- Case studies: how does the model performs in particular episodes, i.e. a recession or an expansion; a period of high or low inflation, etc.
• If models have a Bayesian setup (with proper priors) can use the marginal likelihood. Priors can be non-informative but need to be proper.

Marginal likelihood of model $j$ is $ML(j) = \int f(y, \theta)g(\theta|M(j))d\theta$.

To compare alternatives: Posterior odds ratio/Bayes factor

$$PO = \frac{g(M_j|y)}{g(M_k|y)} = \frac{g(M_j)ML(y|M_j)}{g(M_k)ML(y|M_k)}$$  \hspace{1cm} (12)

The first term is the prior odds, the second the Bayes factor (BF).

Example 2  Want to evaluate the stability of a fixed exchange rate agreement. Under $H_0$ (normal conditions) there is a 50-50 chance that the regime will be maintained (i.e. $f(y = 1|H_0) = f(y = 0|H_0) = 0.5$). Under $H_1$ (say, increasing oil prices) the probability that the fixed exchange
rate regime will be maintained is $0.25 \, f(y = 1|H_0) = 0.75, \, f(y = 0|H_0) = 0.25$. Suppose $g(H_0) = g(H_1) = 0.5$ (equal prior probability), $T = 100$, and that the fixed exchange rate was maintained in 90 periods. Then:

$$PO_{01} = \frac{(0.5)^{0.1}(0.5)^{0.9}}{(0.75)^{0.1}(0.25)^{0.9}} = \frac{0.5}{0.2790} = 1.79$$ (13)

Hence, odds in favor of $H_0$ increased from 1 to 1.79.

- Bayes factors = ratio of marginal likelihood of the two models. It is different than LR statistic!. What matters is the agreement of prior and likelihood and the least square fit of different models. LR does not integrate over $\alpha_j$.

- BF implicitly discounts the fit of large scale models!
• Can also perform posterior predictive analysis (Box, 1980; Faust and Gupta, 2012).

Idea: provide a formal measure of how far a certain feature of the model is at odds with the data

- Can be applied to moments, impulse responses, autocovariances, spectral densities, etc. Only need the features to be a well defined continuous function of the data.

• Canova (1994),(1995) use prior predictive analysis to discard models which are going to be clearly at odds with the data.
Prior predictive analysis: For each $\theta$ simulate a sample $y(\theta)$ from the model, and compute statistics of interest. Can construct distribution of outcomes implied by the model and the prior. Check if the actual value of the statistic is within the range of values produced by the model for that statistic.

- If prior is sufficiently loose, values in the tails indicate that the model should not be used to study that particular phenomena.
Posterior predictive analysis: draw $\theta$ from the posterior and do the exercise as above.

Faust and Gupta (2012): alternative algorithm

- Draw $\theta$ from the posterior, compute $h(Y(\theta))$

- Simulate $Y^d(\theta)$ from every value of $\theta$ you have draws. Compute statistics $h(Y^d(\theta))$

- Plot joint distributions of $h(Y(\theta))$ and $h(Y^d(\theta))$. If they lie around the 45 degree line, data and model agree: otherwise data is unlikely form the point of view of the model.

- The share of points on the 45 degree line is a p-value for the hypothesis that model and data are from the same DGP.

- Apply the technique to SW model.
Figure 3: Posterior density for population, unconditional, correlation of output and inflation (dashed) and posterior predictive density for the sample correlation (solid). Vertical line is the realized value of sample correlation. The numbers in the upper left give the proportion of mass under the posterior and posterior predictive density, respectively, that is to the left of the realized value.
Figure 8: Posterior predictive contour plots for the correlations of various structural shocks. Each panel portrays the joint distribution of $h(Y, \theta | \theta)$ (vertical axis) and $h(Y^*, \theta)$ (horizontal) where $\theta$ is distributed according to the posterior. $Y$ comprises two structural shocks, and $h$ is the sample correlation between the two shocks. The shock labels are: a, productivity; ac, investment productivity; rp, risk premium; pm, price markups; wm, wage markups; g, government spending; mp, monetary policy. The number in the upper left is smaller of the share of points on either side of the 45 degree line.
Figure 9: Posterior predictive contour plots for differences in properties of structural shock in recessions and expansions. Each panel portrays the joint distribution of $h(Y(0), \theta)$ (vertical axis) and $h(Y^e, \theta)$ (horizontal) where $\theta$ is distributed according to the posterior. $Y$ is two smoothed structural shocks, and $h$ is either the absolute difference (recession minus expansion over expansion) in shock standard deviation in the top row, and the simple difference in correlation (recession minus expansion) in the second row. The shock labels are: $a$: productivity; $ai$: investment productivity; $rp$: risk premium; $pm$: price markup; $wm$: wage markup; $g$: government spending; $mp$: monetary policy. The number in the upper left is smaller of the share of points on either side of the 45 degree line.

1) ARMA: $\pi_t = \rho_1 \pi_{t-1} + e_t + \rho_2 e_{t-1}$

2) TVC-BVAR ($\theta = \text{vec}(a_t, b_t(\ell))$):

$$y_t = a_t + b_t(\ell)y_{t-1} + e_t \quad (14)$$
$$\theta_t = \rho \theta_{t-1} + (1 - \rho) \theta_0 + u_t, \quad \theta_t \sim N(0, \Omega) \quad (15)$$
3) NK model:

**IS:** \( x_t = E_t x_{t+1} - \frac{1}{\phi}(r_t - E_t \pi_{t+1}) + g_t \)

**PC:** \( \pi_t = \beta E_t \pi_{t+1} + \frac{\phi(1-\zeta)(1-\beta\zeta)}{\zeta} x_t + u_t \)

**Taylor-Rule:** \( r_t = \psi_r r_t - 1 + (1 - \psi_r) (\phi_x x_{t-1} + \phi_p \pi_{t-1}) + e_t \)

\( v_t = (g_t, u_t) = \rho v_{t-1} + \eta_t; \ \eta_t \ iid \ N(0, \sigma^2). \)

*Pick estimates from Canova (2008)* \( \beta = 0.983(0.0008), \ \phi = 3.04(0.27), \ \zeta = 0.7709(0.185). \)
### In-sample RMSE, percentage points

<table>
<thead>
<tr>
<th>Model</th>
<th>ARIMA</th>
<th>BVAR-TVC</th>
<th>NK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.88</td>
<td>1.04</td>
<td>1.33</td>
</tr>
</tbody>
</table>

### In-sample, correlations: actual and predicted

<table>
<thead>
<tr>
<th>Model</th>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>0.67</td>
<td>0.88</td>
<td>0.76</td>
</tr>
<tr>
<td>BVAR-TVC</td>
<td>0.77</td>
<td>0.89</td>
<td>0.72</td>
</tr>
<tr>
<td>NK</td>
<td>0.56</td>
<td>0.68</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Why is the MSE different from in-sample correlation analysis?

- MSE sum of square bias and variance. MSE could be small is variance small and bias not too large - a "straight line prediction" (a random walk) is typically good.

- Good MSE does not mean that actual and predicted go up and down together.

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
<th>p-value $a = 0, b = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>0.159 (2.01)</td>
<td>0.79 (1.88)</td>
<td>0.03</td>
</tr>
<tr>
<td>BVAR-TVC</td>
<td>0.109 (1.56)</td>
<td>0.67 (2.06)</td>
<td>0.02</td>
</tr>
<tr>
<td>NK</td>
<td>0.035 (0.99)</td>
<td>0.56 (1.71)</td>
<td>0.01</td>
</tr>
</tbody>
</table>
### Predictive regressions

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA-NK</td>
<td>0.82 (2.17)</td>
<td>0.23 (1.65)</td>
</tr>
<tr>
<td>BVAR-NK</td>
<td>0.73 (1.96)</td>
<td>0.35 (2.00)</td>
</tr>
</tbody>
</table>

### Output growth VAR regressions

<table>
<thead>
<tr>
<th>Model</th>
<th>p-value lags of predicted inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just lagged output</td>
<td>0.00</td>
</tr>
<tr>
<td>Adding Inflation</td>
<td>0.15</td>
</tr>
<tr>
<td>Adding Nominal Rate</td>
<td>0.42</td>
</tr>
</tbody>
</table>

- Bayes factor: $\text{ML}(\text{NK})/\text{ML}(\text{BVAR}) = 0.02$. 
2.2.2 Economic tests

- Compute moments of actual data and predicted ones.
- Compute dynamic responses to shocks in the model and the data.
- Compute turning points statistics, overall or at some dates.
- Compare favorite stylized facts with model implications.
<table>
<thead>
<tr>
<th>Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>ARIMA</td>
</tr>
<tr>
<td>BVAR</td>
</tr>
<tr>
<td>NK</td>
</tr>
<tr>
<td>Actual</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Peak inflation, late 1970s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Actual</td>
</tr>
</tbody>
</table>
2.3 Forecasting with DSGE models

Recall: Log linearized aggregate decision rule of a DSGE model is:

\[
\begin{align*}
y_{2t} &= A_{22}(\theta)y_{2t-1} + A_{21}(\theta)y_{3t} \\
y_{1t} &= A_{1}(\theta)y_{2t} = A_{11}(\theta)y_{2t-1} + A_{12}(\theta)y_{3t}
\end{align*}
\]

(16) (17)

\(y_{2t}\) = states and the driving forces, \(y_{1t}\) = controls, \(y_{3t}\) shocks. \(A_{i,j}(\theta), i, j = 1, 2\) are time invariant functions of \(\theta\), the structural parameters.

- There are cross equation restrictions since \(\theta_i, i = 1, \ldots, n\) appears in more than one entry of these matrices.

- (17) is a state space or a restricted VAR(1) model
Unconditional forecast:  $y_{3t+\tau} = 0, \forall \tau > 0$, let the system run. With a VAR(1) representation: let $y_t = (y_{1t}, y_{2t})$. Then $y_{t+\tau} = \hat{A}^\tau y_t$ and $y_{2t+\tau} = S\hat{A}^\tau$, where $\hat{A}$ is an estimate of $A$ and $S$ is a selection matrix, picking up the second set of elements from $A$.

To calculate uncertainty around point forecasts.
If a distribution for $\hat{A}$ is available (asymptotic or posterior) then:

1. Draw $A^l$ from this distribution, compute $y_{l+\tau}^l$, $l = 1, 2, \ldots, L$, each horizon $\tau$.

2. Order $y_{l+\tau}^l$ over $l$, each $\tau$ and extract 16-84 or 2.5-97.5 percentiles.
Conditional forecast 1: Manipulating shocks.

This is the same as computing impulse responses, i.e. need to orthogonalize the disturbances if they are not orthogonal. Only difference is that here impulse may last more than one period. Choose $y_{3t+\tau} = \bar{y}_{3t+\tau}$, $\tau = 0, 1, 2, \ldots, \bar{\tau}$. Given $\hat{A}$ find $y_{2t+\tau} = \hat{A}_{22}(\theta)y_{2t+\tau-1} + \hat{A}_{21}(\theta)y_{3t+\tau}$ and $y_{1t+\tau} = \hat{A}_{1}(\theta)y_{2t+\tau}$.

To calculate uncertainty around the forecasted path, use same algorithm employed for unconditional forecasts (i.e. draw $A$’s from their distributions).
Conditional Forecast 2: Manipulating endogenous states

This requires backing out shocks needed to produce the path $\bar{y}_{2t+\tau}$, $\tau = 0, 1, 2, \ldots$. Simply use the first equation of (17) to do this. Then $y_{1t+\tau} = A_1(\theta) \bar{y}_{2t+\tau}$, $\tau = 1, 2, \ldots$. Same as above to compute uncertainty around the forecasted path.

Identification problem: there may be different elements of $y_{3t}$ which may induce the require path for $y_{2t+\tau}$.

**Example 4** What is the range of paths for consumption from next quarter up to 10 years if the capital stock is higher by ten percent in all these periods? Question: how do we increase the capital stock? Via technology shocks? Via labor supply shocks?
• Conditional Forecast 3: Manipulating endogenous controls. Separate $y_{1t} = [y_{1t}^A, y_{1t}^B]$ and $y_{1t+\tau}^A = \bar{y}_{1t+\tau}^A$, $\tau = 0, 1, 2, ...$. Back out the path of $y_{1t+\tau}^A$ needed to produce $\bar{y}_{1t+\tau}^A$. With this path compute $y_{1t+\tau}^B$. Same identification problems as above; less problematic in some cases.

**Example 5** Suppose that interest rates are (discretionarily) kept 50 basis point higher than the endogenous Taylor rule would imply. What is the effect on inflation?
2.3.1 Out-of-sample evaluation

Use the same statistics employed for in-sample analysis. Now can evaluate forecasts at different horizons.

Out-of-sample RMSE, percentage points, unconditional forecasts

<table>
<thead>
<tr>
<th>Model</th>
<th>1 quarter</th>
<th>4 quarters</th>
<th>8 quarters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>1.43</td>
<td>2.16</td>
<td>2.92</td>
</tr>
<tr>
<td>BVAR-TVC</td>
<td>1.21</td>
<td>1.72</td>
<td>1.89</td>
</tr>
<tr>
<td>NK</td>
<td>1.33</td>
<td>1.58</td>
<td>1.87</td>
</tr>
</tbody>
</table>

Unconditional Out-of-sample Predictive regressions, estimates of $b$

<table>
<thead>
<tr>
<th>Model</th>
<th>1 quarter</th>
<th>4 quarters</th>
<th>8 quarters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA-NK</td>
<td>0.35 (1.71)</td>
<td>0.42 (1.97)</td>
<td>0.34 (2.00)</td>
</tr>
<tr>
<td>BVAR-NK</td>
<td>0.17 (1.66)</td>
<td>0.35 (1.89)</td>
<td>0.44 (2.06)</td>
</tr>
</tbody>
</table>
3 Exploiting other features for evaluation

- Distinguish between internal vs. external dynamics. Models driven almost entirely by external dynamics not very useful.

- Plot together shocks (not the residuals) and the data. What is the contribution of the model?

- Co-cycles analysis.

Log-linearized model (no distinction now states/controls):

\[ y_t = B(\theta)E_t y_{t+1} + A(\theta)y_{t-1} + F(\theta)u_t \]  

(18)
Solution:

\[ y_t = Py_{t-1} + D \sum_{j} S^j E_t u_{t+j} \]  \hspace{1cm} (19)

where \( P - BP^2 - A = 0, \ D = (I - BP)^{-1}F, \ S = (I - BP)^{-1}B. \) If \( u_t = \Phi u_{t-1} + \eta_t \) and we let \( G = D \sum_{j} S^j \Phi j \)

\[ y_t = Py_{t-1} + Gu_t = Py_{t-1} + G\Phi u_{t-1} + G\eta_t \]  \hspace{1cm} (20)
Interested in features of $P$ and $G$.

If $\text{rank}(G) \leq \text{dim}(u)$, let $G^+ = (G'G')^{-1}G'$ and $u_t = G^+(y_t - Py_{t-1})$ so that

$$
y_t = Py_{t-1} + G\Phi(G^{-1}(y_{t-1} - Py_{t-2}) + G\eta_t
= (P + G\Phi G^{-1})y_{t-1} + G\Phi G^{-1}Py_{t-2} + G\eta_t
$$

(21) (22)

Since $\text{rank}(P) = \min(\text{rank}(I - BP), \text{rank}(A))$, if $A$ is of reduced rank (i.e. more states than controls), $P$ will be of reduced rank, i.e. comovements in $y_t$ driven by a reduced number of shocks. (Note this is different from the fact that $\text{dim}(u) < \text{dim}(y)$).

Is this true in the data? Can use factor models to verify this.
• When models feature both long run and short run dynamics Possibility of evaluation looking at long run features.

- Check if permanent component of the model has the same properties as permanent component in the data.

- Can use both cointegration or BQ decompositions. If cointegration: $\Delta y_t = C(\ell) \eta_t$. Interested in $C(1)$. Choose $\beta' C(1) = 0$. Partition $y_t = [y_{1t}, y_{2t}]$ where $y_{2t}$ are I(0). Then $\phi_t = \beta' y_{1t}$ are cointegrating vectors and the VECM is

$$
\begin{bmatrix}
\Delta y_{1t} \\
y_{2t}
\end{bmatrix} = \begin{bmatrix}
\alpha' \\
\gamma'
\end{bmatrix} \phi_{t-1} + My_{2t-1} + v_t
$$

Use this to extract the permanent component of $y_{1t}$ in model and data and to compare them.
Blanchard-Quah decomposition is:

\[
\begin{pmatrix}
\Delta y_{1t} \\
\Delta y_{2t}
\end{pmatrix} = \begin{pmatrix}
\bar{y}_1 \\
0
\end{pmatrix} + \begin{pmatrix}
C_1(1) \\
0
\end{pmatrix} e_t + \begin{pmatrix}
(1 - \ell)C_1^+(\ell) \\
(1 - \ell)C_2^+(\ell)
\end{pmatrix} e_t
\]

where \( C_1^+(\ell) = \frac{C_1(\ell) - C_1(1)}{1 - \ell} \quad C_2^+(\ell) = \frac{C_2(\ell)}{1 - \ell} \), \( 0 < \text{rank}[C_1(1)] \leq m_1 \) and \( \Delta y_t^\infty = [\bar{y}_1 + C_1(1)e_t, 0]' \) is the permanent component of \( y_t \).

- May want to choose (estimate) parameters so that the model tracks long run properties of the data. Then check if dynamics (univariate) properties fit.
A few issues to keep in mind

- Estimated DSGE models typically have driving forces that are correlated (theory assumes that they are not). Misspecification!!!

- If the number of variables is different from the number of shocks, solution is not a VAR but a VARMA: invertibility problems!

- Common to add measurement error but be careful:

\[ y_t = y_t^* + e_t \]  \hspace{1cm} (24)

\[ y_t^* = Py_{t-1}^* + Gu_t \]  \hspace{1cm} (25)

a) if \( e_t \) is iid → signal extraction problem, use Kalman Filter to get \( y_t^* \).
b) If $e_t$ is serially correlated ($e_t = \rho e_{t-1} + v_t$) then:

$$\Delta y_t = \Delta y_t^* + (1 - \rho)(y_{t-1}^* - y_{t-1}) + v_t$$

(26)

This is VECM linking observables $y_t$ and unobservables $y_t^*$.

- On average $y_t = y_t^*$. In short run deviations are possible.

- Can’t use the KF to construct $y_t^*$ in this case.
4 Evaluation via VARs

- Long history in the literature

Canova, Finn, Pagan (1993): Evaluate quantitative properties of RBC models through VARs.

Ingram and Whiteman (1994): Use model to setup a prior for the VAR. Is it better than standard statistical priors?

Canova and Paustian (2007): Evaluate model using qualitative model-based sign restrictions to identify shocks in a VAR.
Procedure

- Start with a broad class of structural models. The class should nest submodels through parameter restrictions (price and wage stickiness, indexation, habit,...).

- Find implications that are robust to parameter variations.

  (a) Some implications are robust across submodels.

  (b) Some implications are robust within a particular submodel.

- Use a subset of implications robust across models to identify shocks in a VAR.

- Use implications that are robust within a submodel and different across submodels for evaluation.

- Do this qualitatively and quantitatively using probabilistic criteria.
Details


- Robust testing: sign and shape of dynamics of unrestricted variables to shocks.

- Produce a partially identified model: standard statistical criteria problematic (Moon and Schorfheide (2008)).

- Can be used **without** estimation of the parameters - good if there are big identification problems.

- VAR mispecification (relative to a DSGE) ok.
Why VAR misspecification not a problem?

- Use robust sign restriction.

- Shock identification robust to time series representation of decision rules.

\[
x_{1t} = A(\theta)x_{1t-1} + B(\theta)e_t \\
x_{2t} = C(\theta)x_{1t-1} + D(\theta)e_t
\]  

(27)
\[
\begin{bmatrix}
I - F_{11}\ell & F_{12}\ell \\
F_{21}\ell & I - F_{22}\ell
\end{bmatrix}
\begin{bmatrix}
y_{1t} \\
y_{2t}
\end{bmatrix}
= \begin{bmatrix}
G_1 \\
G_2
\end{bmatrix} e_t
\]

Representation for \(y_{2t}\) (integrating out \(y_{1t}\)):

\[
(I - F_{22}\ell - F_{21}F_{12}(1 - F_{11}\ell)^{-1}\ell^2) y_{2t} = [G_2 - (F_{21}(1 - F_{11}\ell)^{-1}G_1\ell)] e_t
\]

(28)

ARMA(\(\infty, \infty\)) but impact effects of \(e_t\) has correct sign and magnitude.
Example 6 (use model based restrictions to robustify inference). Use Christiano, et. al. (2005) and Smets and Wouters (2003) class of models.

\[ y_t = c_y c_t + i_y i_t + g_y e_t^g \]  \hspace{1cm} (29)

\[ c_t = \frac{h}{1 + h} c_{t-1} + \frac{1}{1 + h} E_t c_{t+1} - \frac{1 - h}{(1 + h)\sigma_c} (R_t - E_t \pi_{t+1}) + \frac{1 - h}{(1 + h)\sigma_c} (e_t^b - E_t e_{t+1}^b) \]  \hspace{1cm} (30)

\[ i_t = \frac{1}{1 + \beta} i_{t-1} + \frac{\beta}{1 + \beta} E_t i_{t+1} + \frac{\phi}{1 + \beta} q_t - \frac{\beta E_t e_{t+1}^I - e_t^I}{1 + \beta} \]  \hspace{1cm} (31)

\[ q_t = \beta (1 - \delta) E_t q_{t+1} - (R_t - \pi_{t+1}) + \beta r^* E_t r_{t+1} \]  \hspace{1cm} (32)

\[ y_t = \omega (\alpha K_{t-1} + \alpha \psi r_t + (1 - \alpha) l_t + e_t^r) \]  \hspace{1cm} (33)

\[ k_t = (1 - \delta) k_{t-1} + \delta i_t \]  \hspace{1cm} (34)

\[ \pi_t = \frac{\beta}{1 + \beta \mu_p} E_t \pi_{t+1} + \frac{\mu_p}{1 + \beta \mu_p} \pi_{t-1} + \kappa_p m c_t \]  \hspace{1cm} (35)

\[ w_t = \frac{\beta}{1 + \beta} E_t w_{t+1} + \frac{1}{1 + \beta} w_{t-1} + \frac{\beta}{1 + \beta} E_t \pi_{t+1} - \frac{1 + \beta \mu_w}{1 + \beta} \pi_t + \frac{\mu_w}{1 + \beta} \pi_{t-1} - \kappa_w \mu_t^W \]  \hspace{1cm} (36)

\[ l_t = -w_t + (1 + \psi) r_t + k_{t-1} \]  \hspace{1cm} (37)

\[ R_t = \rho_R R_{t-1} + (1 - \rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^R \]  \hspace{1cm} (38)
### Support for the parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_c$</td>
<td>risk aversion coefficient</td>
</tr>
<tr>
<td>$h$</td>
<td>consumption habit</td>
</tr>
<tr>
<td>$\sigma_l$</td>
<td>inverse labor supply elasticity</td>
</tr>
<tr>
<td>$\omega$</td>
<td>fixed cost</td>
</tr>
<tr>
<td>$1/\phi$</td>
<td>adjustment cost parameter</td>
</tr>
<tr>
<td>$\delta$</td>
<td>capital depreciation rate</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>capital share</td>
</tr>
<tr>
<td>$1/\psi$</td>
<td>capacity utilization elasticity</td>
</tr>
<tr>
<td>$g_y$</td>
<td>share of government consumption</td>
</tr>
<tr>
<td>$\zeta_p$</td>
<td>degree of price stickiness</td>
</tr>
<tr>
<td>$\mu_p$</td>
<td>price indexation</td>
</tr>
<tr>
<td>$\zeta_w$</td>
<td>degree of wage stickiness</td>
</tr>
<tr>
<td>$\mu_w$</td>
<td>wage indexation</td>
</tr>
<tr>
<td>$\varepsilon^w$</td>
<td>steady state markup in labor market</td>
</tr>
<tr>
<td>$\gamma_R$</td>
<td>lagged interest rate coefficient</td>
</tr>
<tr>
<td>$\gamma_\pi$</td>
<td>inflation coefficient on interest rate rule</td>
</tr>
<tr>
<td>$\rho_y$</td>
<td>output coefficient on interest rate rule</td>
</tr>
<tr>
<td>$\varrho_i$</td>
<td>persistence of shocks $i = 1, \ldots, 7$</td>
</tr>
</tbody>
</table>
Question of interest: What is the relationship between hours and technology shocks? Do hours robustly fall or robustly increase?

Sign of the impact responses to shocks

<table>
<thead>
<tr>
<th></th>
<th>TFP</th>
<th>Monetary</th>
<th>Taste</th>
<th>Inv</th>
<th>Markup</th>
<th>L^e</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δy_t</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>π_t</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Δc_t</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Δgap_t</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Δw_t</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>?</td>
</tr>
</tbody>
</table>
Identification restrictions for technology shocks

a) $\pi \downarrow, \Delta y \uparrow$.

b) $\Delta c \downarrow$ with Investment shock, $\uparrow$ with others.

c) $\Delta gap \uparrow$ with TFP shocks, $\downarrow$ with markup

d) $\Delta w \downarrow$ with Labor supply and investment shocks, $\uparrow$ markup and TPF shocks.

- These restrictions produce mutually exclusive shocks.

- These restrictions do not involve hours. Once shocks are identified measure the response of hours and the contribution of various technology shocks to their variability.
Responses of hours to technology shocks
Share of hours volatility explained by technology shocks
Conclusions: a model based identification approach tells us:

- Response of hours depend on the source of technological disturbance

- With TFP shocks hours response insignificant contemporaneously, weakly positive after a while (i.e. it is neither NK not RBC).

- With the other shocks hours typically increase.

- Proportion of the variance of hours explained by TPF shocks large but very imprecisely estimated (can’t really say if TPF shocks matter or not).
Example 7 (use model based restrictions test RBC vs. NK transmission) (Pappa, 2009). RBC and NK models have different implications for the transmission of government expenditure shocks to labor markets.

- In RBC a $g$ shock makes hours and wage increase.

- In a NK a $g$ shock makes hours and wage move in the opposite direction.

Which mechanism is more consistent with the facts?

- Take a general specification where you can nest a RBC model as a special case (take a NK model where the monopolistic distortions have been eliminated and consider either a sticky price or a flexible price version of the model).
- Find robust restrictions of the two class of models which do not involve either hours or real wages. There are many. Choose restrictions which are commonly satisfied across models.

<table>
<thead>
<tr>
<th>Shock</th>
<th>Y</th>
<th>Deficit</th>
<th>(C^g)</th>
<th>(I^g)</th>
<th>(N^g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C^g) shock</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td></td>
</tr>
<tr>
<td>(I^g) shock</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td></td>
<td>&gt; 0</td>
<td></td>
</tr>
<tr>
<td>(N^g) shock</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td></td>
<td></td>
<td>&gt; 0</td>
</tr>
</tbody>
</table>

- Contemporaneous effects only (the distinction at longer horizons is blurred).

- Sign consistent with a large range of parameter values.
Aggregate responses
Government consumption shock

Typical responses

Government investment shock

Government employment shock

Figure 3
Labor Market Responses
Del Negro and Schorfheide (2004), Del Negro, et al. (2006): A VAR is a bridge between a DSGE model and the data.

Add to the actual data, simulated DSGE data organized in a VAR. Use a parameter to weight the two information. Model is the DGP of the data if parameter is $\infty$, model totally fails if parameter is 0.

Interpretation: DSGE is a VAR with cross equation restrictions. If restrictions are false better to relax them completely, i.e. use a VAR for the data.
Approach:

- Simulate data from model

- Append simulated data to VAR via a prior (use mean/variance of estimated parameters on simulated data to set up a Normal-Wishart prior).

- Choose proportion of simulated to actual data (to test model).
Let $\theta$ be DSGE parameters; $\alpha$ the VAR parameters. Prior is:

$$g(\theta) = \prod_{i=1}^{k} g(\theta_k);$$

$$g(\alpha) \sim N(\tilde{\alpha}(\theta), \tilde{\Sigma}_a(\theta));$$

$$\Sigma_e \sim IW(T_1 \tilde{\Sigma}(\theta), T_1 - k)$$

where

$$\tilde{\alpha}(\theta) = (X^s'X^s)^{-1}(X^s'y^s)$$

$$\tilde{\Sigma}_a(\theta) = \Sigma_e(\theta) \otimes (T_1 X^s'X^s)^{-1}$$

$$\tilde{\Sigma}(\theta) = (y^s'y^s - (y^s'X^s)\tilde{\alpha}(\theta))$$

(39)

$y^s$ simulated data, $X^s$ lags in the VAR of simulated data. $T_1 =$ length of simulated data. $\kappa = \frac{T_1}{T}$ measures the relative importance of two types of information. $\kappa \to 0$ ($\kappa \to \infty$) actual (simulated) data dominates.
Hierarchical structure: \( f(\alpha, \Sigma_e|y)g(\alpha|\theta)g(\Sigma_e|\theta)g(\theta) \). Since the likelihood and the prior are conjugate:

\[
(\alpha|\theta, y, \Sigma_e) \sim N(\tilde{\alpha}(\theta), \tilde{\Sigma}(\theta));
\]

\[
(\Sigma_e|\theta, y) \sim iW((\kappa + T)\tilde{\Sigma}(\theta), T + \kappa - k)
\]
where

\[
\tilde{\alpha}(\theta) = (T_1X^s'y + X'y)^{-1}(T_1X^s'y + X'y)
\]

\[
\tilde{\Sigma}_a(\theta) = \Sigma_e(\theta) \otimes (T_1X^s'X + X'y)^{-1}
\]

\[
\tilde{\Sigma}(\theta) = \frac{1}{(1 + \kappa)T}[(T_1y^s'y + y'y) - (T_1y^s'X + y'X)\tilde{\alpha}(\theta)]
\]

and \( g(\theta|y) \propto g(\theta) \times |\Sigma_e|^{-0.5(T - M - 1)} \)

\[
\exp\left\{-0.5tr[\Sigma_e^{-1}(Y - X\alpha)'(Y - X\alpha)] \times |\Sigma_e(\theta)|^{-0.5(T_1 - M - 1)}
\exp\left\{-0.5tr[\Sigma_e(\theta)^{-1}(Y^s - X^s\alpha(\theta))'(Y^s - X^s\alpha(\theta))]ight\}.
\]
• Can estimate jointly $\theta$ and $\alpha$ but also possible to calibrate $\theta$.

• All posterior moments in (40) conditional on $\kappa$. How do we select it?
  - Use Rules of thumbs (e.g. $\kappa = 1$, $T$ observation added).
  - Maximize marginal likelihood.
Example 8  In a basic sticky price-sticky wage economy, fix \( \eta = 0.66, \pi^{ss} = 1.005, N^{ss} = 0.33, \frac{c}{gdp} = 0.8, \beta = 0.99, \zeta_p = \zeta_w = 0.75, a_0 = 0, a_1 = 0.5, a_2 = -1.0, a_3 = 0.1. \) Run a VAR with output, interest rates, money and inflation using actual quarterly data from 1973:1 to 1993:4 and data simulated from the model conditional on these parameters. Overall, only a modest amount of simulated data (roughly, 20 data) should be used to set up a prior.

Marginal Likelihood, Sticky price sticky wage model.

<table>
<thead>
<tr>
<th>( \kappa = 0 )</th>
<th>( \kappa = 0.1 )</th>
<th>( \kappa = 0.25 )</th>
<th>( \kappa = 0.5 )</th>
<th>( \kappa = 1 )</th>
<th>( \kappa = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1228.08</td>
<td>-828.51</td>
<td>-693.49</td>
<td>-709.13</td>
<td>-913.51</td>
<td>-1424.61</td>
</tr>
</tbody>
</table>
5 Using loss functions to evaluate DSGE models

Schorfheide (2000): Compare two DSGE models both misspecified.

- PO ratio for misspecified models uninteresting. One model preferred but it may have very close to zero posterior probability.

Example 9 $PO = \frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0} ML(Y_T|M_1)}{\pi_{2,0} f(Y_T)} * \frac{f(Y_T)}{ML(Y_T|M_2)}$. If use 0-1 loss function the posterior risk is minimized by selecting $M_1$ if $PO > 1$.

Potential presence of a third (better specified) model does not affect $PO$ if the prior odds $\frac{\pi_{1,0}}{\pi_{2,0}}$ unchanged ($M_3$ enters only in $f(Y_T)$, which cancels out).

Problem if $M_1$ and $M_2$ have low posterior probability, $M_3$ a large one.
Solution: Use loss functions.

Procedure

1. Compute the posterior distribution for the parameters of each model, using tractable priors and one of the available posterior simulators.

2. Obtain the marginal likelihood of the data, for each $M_i$, that is, compute $f(y|M_i) = \int f(y|\theta_i, M_i)g(\theta_i|M_i)d\theta_i$.

3. Compute posterior probabilities $\tilde{P}_i = \frac{\bar{P}_i f(y|M_i)}{\sum_i \bar{P}_i f(y|M_i)}$, where $\bar{P}_i$ is the prior probability of model $i$. Note that if the distribution of $y$ is degenerated under $M_i$ (e.g. if number of shocks is smaller than the number of endogenous variables), $\tilde{P}_i = 0$. 
4. Calculate the posterior distribution of $h(\theta)$ for each model and average using posterior probabilities i.e. obtain $g(h(\theta)|y, M_i)$, and $g(h(\theta)|y) = \sum_i \tilde{P}_ig(h(\theta)|y, M_i)$. If all but model $i'$ produce degenerate distributions for $\theta$, $g(h(\theta)|y) = g(h(\theta)|y, M_{i'})$.

5. Setup a loss function $\mathcal{L}(h_T, h_i(\theta))$ measuring the discrepancy between model's $i$ predictions and data $h_T$. Since the optimal predictor in model $M_i$ is $\hat{h}_i(\theta) = \arg\min_{h_i(\theta)} \int \mathcal{L}(h_T, h_i(\theta))g(h_i(\theta)|y, M_i)dh_T$, one can compare models using the risk of $\hat{h}_i(\theta)$ under the overall posterior distribution $g(h(\theta)|y)$, i.e. $\mathcal{R}(\hat{h}_i(\theta)|y) = \int \mathcal{L}(h_T, \hat{h}_i(\theta))g(h(\theta)|y)dh_T$. 
In step 5) $\mathcal{R}(\hat{h}_i(\theta) | y)$ measures how well model $\mathcal{M}_i$ predicts $h_T$. Note that while model comparison is relative, $g(h(\theta) | y)$ takes into account information from all models.

Taking step 5) further: for each $i$, $\theta$ can be selected so as to minimize $\mathcal{R}(\hat{h}_i(\theta) | y)$. Such an estimate provides a lower bound to the posterior risk obtained by the "best" candidate model.
• Possible loss functions:

(a) Quadratic loss: \( L_2(h(\theta), \hat{h}(\theta)) = (h(\theta) - \hat{h}(\theta))^\prime W (h(\theta) - \hat{h}(\theta)) \); \( W \) is a weighting matrix.

(b) Penalized Loss: \( L_p(h(\theta), \hat{h}(\theta)) = I[g(h(\theta)|Y_T) > g(\hat{h}(\theta)|Y_T)] \), where \( I(x, z) = 1 \) if \( x > z \) and zero otherwise.

(c) \( \chi^2 \) loss: \( L_{\chi^2}(h(\theta), \hat{h}(\theta)) = I[C_{\chi^2}(h(\theta)|Y_T) > C_{\chi^2}(\hat{h}(\theta)|Y_T)] \) where \( C_{\chi^2}(h(\theta)|Y_T) = (h(\theta) - E(h(\theta)|Y_T))^\prime V_{\theta}^{-1} (h(\theta) - E(h(\theta)|Y_T)) \) and \( V_{\theta} \) is a posterior covariance matrix of \( h(\theta) \).

(d) 0-1 loss: \( L_{01} = 1 \) if \( \hat{h}(\theta) \neq h(\theta) \) and zero otherwise.
• Results:

1) If $g(h(\theta)|Y_T)$ is normal $L_2 = L_{\chi^2}$.

2) Optimal predictor under $L_2$ and $L_{\chi^2}$ is $E(h(\theta)|Y_T, \mathcal{M}_i)$.

3) Optimal predictor for $L_p$ is the posterior mode of $g(h(\theta)|Y_t, \mathcal{M}_i)$.

4) If for any positive definite $W$, $\mathcal{M}_1 > \mathcal{M}_2$ with probability one as $T \to \infty$, $L_q$ selection is consistent and identical to a PO ratio.
5) If the two models are so misspecified that their posterior probability goes to zero as $T \to \infty$, the ranking depends on the discrepancy between $E(h(\theta)|y, \mathcal{M}_3) \approx E(h(\theta)|y)$ and $\hat{h}_i(\theta), i = 1, 2$. If $\mathcal{M}_3$ is any empirical model, then using a $\mathcal{L}_2$ loss is equivalent to compare sample and population moments obtained from different models informally.

Simplest calibration exercises is optimal Bayesian decision using $\mathcal{L}_2$ loss function and the models are highly misspecified.
Example 10 Ferroni (2011).

Take a cyclical DSGE. Possibility that the data is generated by the model plus three alternative specification for the non-cyclical part.

\[ M1 = \text{DSGE} + \text{linear trend}, \quad M2 = \text{DSGE} + \text{HP trend}, \quad M3 = \text{DSGE} + \text{RW trend} \]

1) Which model has the higher posterior probability (starting from an equal prior probability)?

- Log Bayes factor of \( M2 \) relative to \( M1 \) = -31.80.

- Log Bayes factor of \( M3 \) relative to \( M2 \) = 98.47.
2) How do you robustify inference about DSGE parameters to trend uncertainty? Construct

\[ g(\theta|y, DSGE) = \frac{\sum_j p(y|M_j)}{\sum_k p(y|M_k)} \int g(\theta|y, M_j, \alpha^j) d\alpha^j \quad (41) \]

where \( \alpha_j \) represents the parameters of the non-cyclical specification for model \( j \). (Hint to do this you need to estimate cyclical and non-cyclical parts jointly).
6 Some additional thoughts

Different kind of DSGE models:

- Academic model (typically small): generated by the idea of having internal consistency and well defined structure than the need to fit the data.

- Central Bank model (similar to academic model but typically large)

1) Why should a Central Bank model be big??

Can you figure out if it has a unique equilibrium?

Can you figure out what drives dynamics?
Is reality complex or are we unable to understand it?? “Sophisticated simplicity” Jeffreys(1963), Zellner(1981),

2) What defines a Central bank model?

“A tool to help to focus policy discussion around some stories rather than degenerating into a discussion of many special events as it often happens with data driven models”

- Operational Central Bank model (adjusted CB model to fit existing evidence, e.g. adding extra sources of errors or dynamics, converting theoretical variables in measurable ones).

- Central Bank forecasting model (OCB model adjusted to incorporate policymakers beliefs about the future (e.g. incorporate survey data information, information from anticipatory variables, etc.).
• The above evaluation procedures can be applied to any type of CB model. Careful if you use them with last two since posterior information is often used to setup your model.

• How should one move down from theoretical to policy oriented models?? Or should we go the other way around?