Comments on “On the Fit of Forecasting Performance of New-Keynesian Models” by Del Negro, Schorfheide, Smets, and Wouters

James D. Hamilton
UCSD
Data:
1) output
2) consumption
3) investment
4) employment
5) real wage
6) inflation
7) nominal interest rate
Shocks:
1) productivity
2) intertemporal substitution
3) relative cost of investment goods
4) consumption-leisure substitution
5) fiscal shock (fraction to govt)
6) mark-up shock
7) Fed policy
vector of truly structural parameters
(discount rate, serial dependence of productivity shock, capital share,...)

vector of data (growth rates, deviations from mean,...)

cointegrating variables

\[(\ln C_t - \ln Y_t, \ln I_t - \ln Y_t, \ln(W_t/P_t) - \ln Y_t)'\]
DSGE:

\[ y_t = c(\theta) + B(\theta)z_{t-1} + \Phi_1(\theta)y_{t-1} + \Phi_2(\theta)y_{t-2} + \cdots + \Phi_p(\theta)y_{t-p} + \varepsilon_t \]

\[ E(\varepsilon_t\varepsilon_t') = \Omega(\theta) \]

VAR:

\[ y_t = c + Bz_{t-1} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \varepsilon_t \]

\[ E(\varepsilon_t\varepsilon_t') = \Omega \]

\( \theta \) has much fewer elements than VAR
\[ A = \begin{bmatrix} c & B & \Phi_1 & \Phi_2 & \cdots & \Phi_p \end{bmatrix} \]

\[ x_t = (1, z'_t, y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p})' \]

\[ y_t = Ax_t + \varepsilon_t \]

\[ a = \text{vec}(A) \]

\[ a|\theta, \Omega, \lambda \sim N(a(\theta), (\lambda T)^{-1}[\Omega \otimes \Gamma_{xx}(\theta)^{-1}]) \]
\( \mathbf{a|\theta, \Omega, \lambda} \sim N(\mathbf{a(\theta)}, (\lambda T)^{-1}[\Omega \otimes \Gamma_{xx}(\theta)^{-1}]) \)

What is \( \lambda \)?

(1) parameter of prior
bigger \( \lambda \) means more confident in prior info.

If earlier we’d observed a sample of \( \lambda T \) observations for which OLS estimate
was \( \mathbf{a(\theta)} \) and for which we’d had a diffuse
prior, our posterior would have this form
\(a | \theta, \Omega, \lambda \sim N(\mathbf{a}(\theta), (\lambda T)^{-1}[\Omega \otimes \Gamma_{xx}(\theta)^{-1}])\)

\(\lambda\) reflects our confidence in DSGE

\(\lambda = 1\) DSGE counts just as much as current sample

\(\lambda = 2\) DSGE counts twice as much as current sample

\(\lambda = 0\) DSGE counts for nothing
If sample $T$ is twice as big, I have twice as much confidence in the prior.

approximation?

useful rule of thumb?
What is $\lambda$?

(2) hyperparameter

\[ a = a(\theta) + z \]

\[ z \sim \mathcal{N}(0, (\lambda T)^{-1}[\Omega \otimes \Gamma_{xx}(\theta)^{-1}]) \]
Define $\hat{\theta}$ to be the QMLE:

$$\hat{\theta} = \arg \max L(\theta)$$

$$L(\theta) = -(T/2) \ln |\Omega(\theta)|$$

$$- (1/2) \sum_{t=1}^{T} q_t' \Omega(\theta)^{-1} q_t$$

$$q_t = y_t - A(\theta)x_t$$

$$\theta_0 = \text{plim} \hat{\theta}$$

$$A_0 = \text{plim} \hat{A}$$
prior: $a|\theta, \Omega, \lambda \sim$

$$N(a(\theta), (\lambda T)^{-1} \left[ \Omega(\theta) \otimes \Gamma_{xx}(\theta)^{-1} \right])$$

$\lambda$ measures how far we expect $a_0$ to be from $a(\theta_0)$ before seeing data
posterior distributions:

\[ \theta | Y \]
\[ a | Y \]
\[ (\theta, a) | Y \]

\[ \mathbb{E}[(a - a(\theta))(a - a(\theta))'] | Y \]

could calculate ratio of determinant of last magnitude to that of

\[ \mathbb{E}[(a - a_0)(a - a_0)'] | Y \]