The Likelihood Function of a Linearized DSGE Model

Frank Schorfheide
Department of Economics, University of Pennsylvania

Likelihood Function (I)

- Forecast error covariance for vectors of macro variables is typically not rank-deficient.
- In order to explain data, we need a model that delivers a non-degenerate probability distribution for the observables:
  - Include enough structural shocks: e.g. preference shocks. Capture to some extent aggregation effects.
  - Include measurement errors. More precisely, these errors are supposed to capture the discrepancy between model and reality.
- Generalization of shock structure potentially breaks cross-coefficient restrictions. Might reduce misspecification but also introduce additional identification problems.
- Endogenous propagation of innovations versus exogenous propagation.

Likelihood Function (II)

- Likelihood function: joint probability density of the data $Y^T = \{y_1, \dots, y_t\}$:

$$\mathcal{L}(\theta|Y^T) = p(Y^T|\theta) = \prod_{t=1}^{T} p(y_t|Y^{t-1}, \theta) \tag{1}$$

- Log-linearized DSGE models can be written as state-space models:

$$\text{measurement} \;:\; y_t = A(\theta) + B(\theta)s_t \tag{2}$$

$$\text{state transition} \;:\; s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{3}$$

- Make distributional assumption: $\epsilon_t \sim iid\mathcal{N}(0, \Sigma_\epsilon(\theta))$.

- It is only assumed that the y_t's are observable. The vector s_t may have unobservable elements such as conditional expectations or a latent productivity process.

- We obtained the state transition equation when we solved the LRE model.

Likelihood Function (III)

- If s_t is not fully observable we need to use a filter to obtain the likelihood function.

  This filter is a recursive algorithm to calculate

$$p(y_t|Y^{-1},\theta), \quad t=1,\ldots,T$$

- Iterations:

  - Initialization at time t: $p(s_t|Y^t,\theta)$
  - Forecasting $t+1$ given t:
    1. Transition equation: $p(s_{t+1}|Y^t,\theta) = \int p(s_{t+1}|s_t,Y^t,\theta)p(s_t|Y^t,\theta)ds_t$
    2. Measurement equation: $p(y_{t+1}|Y^t,\theta) = \int p(y_{t+1}|s_{t+1},Y^t,\theta)p(s_{t+1}|Y^t,\theta)ds_{t+1}$
  - Updating with Bayes theorem. Once y_{t+1} becomes available:

$$p(s_{t+1}|Y^{t+1},\theta) = p(s_{t+1}|y_{t+1},Y^t,\theta) = \frac{p(y_{t+1}|s_{t+1},Y^t,\theta)p(s_{t+1}|Y^t,\theta)}{p(y_{t+1}|Y^t,\theta)}$$

Likelihood Function (IV)

- Initialization: process s_t is stationary we can initialize the filter with the unconditional distribution of s_t, calculated from the transition equation.

$$\mathbb{E}[s_t s_t'] = \Phi_1 \mathbb{E}[s_t s_t'] \Phi_1' + \Phi_\epsilon \Sigma_\epsilon \Phi_\epsilon'$$

- Iterations look difficult because they involve integrations. But:
- If ϵ_t is normally distributed, all conditional distributions are also normal.
- At each step we only track means and covariance matrices... Kalman filter.
- Kalman filter iterations can be found in standard time series textbooks, e.g., Hamilton (1994). (see Appendix).

Kalman Filter

- I am using slightly different notation now...

- **Measurement equation**: $y_t = A + Bs_t + u_t$,

  u_t's are innovations (or "measurement errors") with mean zero and $\mathbb{E}_{t-1}[u_t u_t'] = H$.

- **Transition equation**: $s_t = \Phi s_{t-1} + R\epsilon_t$,

  where ϵ_t is a vector of innovations with mean zero and variance Σ_ϵ.

- System matrices $A, B, \Phi, R, \Sigma_\epsilon, H$ are non-stochastic and predetermined, so system is linear and y_t can be expressed as a function of present and past u_t's and ϵ_t's.

- Write $\mathbb{E}[s_0] = S_0$ and $Var[s_0] = P_0$.

Kalman Filter

- **Initialization**: Start with prior distribution for the initial state s_0: $s_0 \sim \mathcal{N}(S_0, P_0)$, e.g., choose S_0 and P_0 to be mean and variance of (stationary) state vector distribution.

- **Forecasting**: At $(t-1)^+$, that is, after observing y_{t-1}, the belief about the state vector has the form $s_{t-1}|Y^{t-1} \sim \mathcal{N}(S_{t-1}, P_{t-1})$. Thus, the "posterior" from period $t-1$ turns into a prior for $(t-1)^+$.

- Since s_{t-1} and ϵ_t are independent multivariate normal random variables, it follows that

$$s_t|Y^{t-1} \sim \mathcal{N}(\hat{s}_{t|t-1}, P_{t|t-1}) \tag{4}$$

where

$$\begin{aligned} \hat{s}_{t|t-1} &= \Phi S_{t-1} \\ P_{t|t-1} &= \Phi P_{t-1} \Phi' + R \Sigma_\epsilon R' \end{aligned}$$

Kalman Filter

- The conditional distribution of $y_t|s_t, Y^{t-1}$ is of the form

$$y_t|s_t, Y^{t-1} \sim \mathcal{N}(A + Bs_t, H) \tag{5}$$

Since $s_t|Y^{t-1} \sim \mathcal{N}(\hat{s}_{t|t-1}, P_{t|t-1})$, we can deduce that the marginal distribution of y_t conditional on Y^{t-1} is of the form

$$y_t|Y^{t-1} \sim \mathcal{N}(\hat{y}_{t|t-1}, F_{t|t-1}) \tag{6}$$

where

$$\begin{aligned} \hat{y}_{t|t-1} &= A + B\hat{s}_{t|t-1} \\ F_{t|t-1} &= BP_{t|t-1}B' + H \end{aligned}$$

Kalman Filter

- **Updating**: To obtain the posterior distribution of $s_t|y_t, Y^{t-1}$ note that

$$\begin{aligned} s_t &= \hat{s}_{t|t-1} + (s_t - \hat{s}_{t|t-1}) && (7) \\ y_t &= \hat{y}_{t|t-1} + B(s_t - \hat{s}_{t|t-1}) + u_t && (8) \end{aligned}$$

- and the joint distribution of s_t and y_t is given by

$$\left[\begin{array}{c} s_t \\ y_t \end{array} \right] \Big| Y^{t-1} \sim \mathcal{N}\left(\left[\begin{array}{c} \hat{s}_{t|t-1} \\ \hat{y}_{t|t-1} \end{array} \right], \left[\begin{array}{cc} P_{t|t-1} & P_{t|t-1}B' \\ BP'_{t|t-1} & F_{t|t-1} \end{array} \right] \right) \quad (9)$$

Kalman Filter

- Applying Bayes theorem, i.e., calculating a conditional distribution based on a joint...

$$s_t | y_t, Y^{t-1} \sim \mathcal{N}(S_t, P_t) \tag{10}$$

where

$$\begin{aligned} S_t &= \hat{s}_{t|t-1} + P_{t|t-1} B' F_{t|t-1}^{-1} (y_t - A - B\hat{s}_{t|t-1}) \\ P_t &= P_{t|t-1} - P_{t|t-1} B' F_{t|t-1}^{-1} B P_{t|t-1} \end{aligned}$$

The conditional mean and variance $\hat{y}_{t|t-1}$ and $F_{t|t-1}$ were given above. This completes one iteration of the algorithm. The posterior $s_t|Y^t$ is the prior for the next iteration.

Kalman Filter

- **Likelihood Function**:

$$
\begin{aligned}
p(Y^T|\text{parameters}) &= \prod_{t=1}^{T} p(y_t|Y^{t-1}, \text{parameters}) \qquad (11) \\
&= (2\pi)^{-nT/2} \left(\prod_{t=1}^{T} |F_{t|t-1}| \right)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \nu_t F_{t|t-1} \nu_t' \right\}
\end{aligned}
$$

where

$$
\nu_t = y_t - \hat{y}_{t|t-1} = B(s_t - \hat{s}_{t|t-1}) + u_t
$$

Extensions – Conditioning (I)

- Conditioning: so far we discussed the unconditional likelihood function.

- Other time series models, such as VARs, are commonly estimated based on conditional likelihood functions, e.g.,

$$p(y_1, \ldots, y_T | y_0, \theta) \tag{12}$$

- In order to obtain a conditional likelihood for the DSGE, note that

$$p(y_1, \ldots, y_T | y_0, \theta) = \frac{p(y_0, y_1, \ldots, y_T | \theta)}{p(y_0 | \theta)} \tag{13}$$

- The numerator can be obtained by applying the (Kalman) filter to all observations, including y_0, whereas the denominator is obtained by applying the filter to y_0 only.

Extensions – Conditioning (II)

- Short-cut: the Kalman filter applied to $y_0, \ldots, y_T$ generates $p(y_0)$ and $p(y_t|Y^{-1})$, for $t = 1, \ldots, T$. Calculate conditional likelihood as

$$\mathcal{L} = \prod_{t=1}^{T} p(y_t|Y^{t-1})$$

Extensions – Nonstat Levels (I)

- Reference: Chang, Doh, Schorfheide (2005): "Non-stationary Hours in a DSGE Model."

- Log-linearized DSGE has state space representation of the form:

$$
\begin{aligned}
y_t &= \Gamma_0 + \Gamma_1 s_{1,t} + \Gamma_2 s_{2,t} + \Gamma_3 t && (14)\\
s_{1,t} &= \Phi_1 s_{1,t-1} + \Psi_1 \epsilon_t && (15)\\
s_{2,t} &= s_{2,t-1} + \Psi_2 \epsilon_t. && (16)
\end{aligned}
$$

- Trend in (14) captures drift random walk technology A_t.

- Equation (15) represents the law of motion for the variables of the detrended model.

- (16) describes evolution of $s_{2,t} = \ln A_t - \gamma t$ for $\mathcal{M}_0$ and $s_{2,t} = [\ln A_t - \gamma t, \ln B_t]'$.

Extensions – Nonstat Levels (II)

- Initialization of Kalman filter: factorize the initial distribution as $p(s_{1,0})p(s_{2,0})$.

- Set the first component equal to the unconditional distribution of $s_{1,t}$.

- Absorb second component, composed of the distribution of $\ln A_0$ ($\mathcal{M}_0$) and $[\ln A_0, \ln B_0]'$ ($\mathcal{M}_1$), respectively, into the specification of our prior $p(\theta)$.

- If DSGE model implies, say, common trend in investment, consumption, and output, then estimation based on likelihood function in levels, incorporates information about the ratios investment/output, consumption/output.

- Roughly, likelihood will relate the model implied steady-state ratios to long-run averages in the data.

Likelihood Function (V)

- If DSGE models is solved with nonlinear method then Kalman Filter is not sufficient.
- Alternative: particle filter or sequential Monte Carlo filter.
- Particle has been used to analyze stochastic volatility models: Pitt and Shephard (1999), and Kim, Shephard, and Chib (1998);
- DSGE models solved with finite elements method: Fernandez-Villaverde and Rubio-Ramirez (2004).
- DSGE models solved with second-order perturbation methods: An (2005) and An and Schorfheide (2005).