Lecture 1: Estimating New Keynesian DSGE Models via Linearization & Bayesian Methods

James M. Nason

Department of Economics
North Carolina State University
February 13, 2018

What Makes a Macro Model Keynesian

- Since Keynes (1964, The General Theory of Employment, Interest, and Prices, New York, NY: Harcourt, Brace, and World, Inc.), Keynesian economics argues classical theory (*i.e.*, all prices fully flexible) is valid only at the full employment level of output.

- The important point to understand is that Keynes stood classical theory on its head.
  1. Keynes altered the focus of (what we call macroeconomics) from the determination of the price level to the determination of output (or more accurately, aggregate demand).
  2. Keynes accomplished this by assuming (implicitly) some price somewhere does not adjust (*i.e.* is sticky) in the manner in which classical theory predicts it should at the relevant business cycle horizon.
  3. This assumption remains the fundamental debating point between economists operating in the classical and Keynesian traditions today.
  4. See Tobin (1980, Asset Accumulation and Economic Activity, Chicago, IL: University of Chicago Press), Summers (1986, "Some skeptical observations on real business cycle theory," *Quarterly Review*, Federal Reserve Bank of Minneapolis 10, 23–29), and
  5. Prescott (1986, "Theory ahead of business cycle measurement," *Quarterly Review*, Federal Reserve Bank of Minneapolis 10, 9–22).

What Makes a Keynesian Model New Keynesian

- Keynes (1964) and Summers (1986) are consistent with New Keynesian (NK) models.
  1. Aggregate demand determination is at the heart of NK models.
  2. Monopolistically competitive firms (households) supply output (labor services) to meet demand, which is downward sloping.
  3. $\Longrightarrow$ These firms (households) control the prices (nominal wages) of the goods (labor services) they sell into the goods (labor) market.

- NK models are draped on the edifice of real business cycle (RBC) theory having these features $\Rightarrow$ start with a complete markets model; a good example is Devereux, Head, and Lapham (1996, "Aggregate fluctuations with increasing returns to specialization and scale," *Journal of Economic Dynamics and Control* 20, 627–656).

- Discipline NK nominal stickiness with a story having micro-foundations $\Longrightarrow$ Calvo (1983, "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics* 12, 383–398); for an alternative approach see Sims (1998, "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49, 317–356).
  1. Calvo staggered price (nominal wage) setting exists because only a fraction $1 - \mu_P$ $(1 - \mu_W)$ of the monopolistically competitive final goods firms (households) are able to set and commit to a new output price, $P_{C,t}$ (nominal wage, $W_{C,t}$), between date $t-1$ and date $t \Longrightarrow$ time-dependent staggered price setting.
  2. The Calvo staggered price setting story is first integrated into a monetary DSGE by Yun (1996, "Nominal rigidities, money supply endogeneity, and business cycles," *Journal of Monetary Economics* 37, 345–370).

Introduction to NKDSGE Models

- Consider a NKDSGE model that has households, firms, a government, and feature
  1. real frictions: preferences display internal habit formation in consumption and varying capital utilization and capital or investment adjustment are costly,
  2. nominal frictions: prices and nominal wages are assumed sticky à la Calvo,
  3. the economy grows along a stochastic balanced growth path created by TFP, which is random walk (with drift) $\Longrightarrow$ the stochastic trend of the economy,
  4. and a monetary authority endowed with a money growth or Taylor rule.
  5. NKDSGE models lack complete markets because of the nominal stickiness not these real frictions.

- This medium scale NKDSGE model is similar to those estimated by
  1. Del Negro and Schorfheide (2008, "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *JME* 55, 1191–1208),
  2. Del Negro, Schorfheide, Smets, and Wouters (2007, "On the fit and forecasting performance of new Keynesian models," *JBES* 25, 123–162),
  3. Smets and Wouters (2007, "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *AER* 97, 586–606),
  4. CEE (2005, "Nominal rigidities and the dynamic effects of a shock to monetary policy," *JPE* 113, 1–45),
  5. Smets and Wouters (2003, "An estimated stochastic dynamic general equilibrium model of the Euro area," *JEEA* 1, 1123–1175).
  6. The dynamic responses of output and inflation to a monetary policy shock mark the empirical success of NKDSGE models $\Longrightarrow$ few if any other classes of DSGE models generate similar responses.
  7. All except CEE add exogenous (*i.e.*, preference, markup, and/or government spending) shocks, which are often assumed to be AR(1)s processes.

Primitives: The Demand for Goods and Technology

- A continuum of monopolistically competitive firms produce final goods households consume $\Longrightarrow$ firms and their differentiated goods have addresses $j \in [0, 1]$.

1. The consumption aggregator is $c_t = \left[\int_0^1 y_{D,t}(j)^{\frac{\xi-1}{\xi}} dj\right]^{\frac{\xi}{\xi-1}}$, where the price elasticity is $\xi > 1$ and $y_{D,t}(j)$ is household final good demand for the output of firm j; see Dixit and Stiglitz (1977, "Monopolistic competition and optimum product diversity," *American Economic Review* 67, 297–308).
2. The jth final good firm aims to meet this demand with its output, $y_t(j)$, by mixing efficiency units of capital, $u_t K_t(j)$, rented and labor, $N_t(j)$, hired from households net of fixed labor costs, N_0, given labor-augmenting TFP, A_t, in the CRS technology,

$$y_t(j) \;=\; \Big[u_t K_t(j)\Big]^{\psi}\Big[\big(N_t(j) - N_0\big) A_t\Big]^{1-\psi}, \quad \psi \in (0, 1),$$

3. where u_t = capital utilization rate, $u_t \in (0, 1]$, and a fixed labor cost $N_0 > 0$, which is necessary for monopolistic competition in the final goods market.
4. For the NKDSGE model to have a permanent shock, TFP is a random walk with drift, $A_t = A_{t-1}\exp\{\alpha + \varepsilon_t\}$, with its innovation, $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

Primitives: Staggered Price Setting of Output, I

- Firm j maximizes profits by setting its price $P_t(j)$, s.t. $Y_{D,t} = \left[\int_0^1 y_{D,t}(j)^{\frac{\xi-1}{\xi}} dj\right]^{\frac{\xi}{\xi-1}}$

  $$\Longrightarrow \ \textbf{Max}_{\,y_{D,t}(j)} \ \int_0^1 P_t(j) y_{D,t}(j) dj \ + \ P_t \left[Y_{D,t} - \left[\int_0^1 y_{D,t}(j)^{\frac{\xi-1}{\xi}} dj\right]^{\frac{\xi}{\xi-1}} \right],$$

  where $Y_{D,t} \equiv c_t$ is another way to express aggregate demand and P_t is the aggregate price level, which firm j takes as given.

- The solution is the downward sloping demand schedule $y_{D,t}(j) = \left[\frac{P_t}{P_t(j)}\right]^{\xi} Y_{D,t}$ and the price index $P_t = \left[\int_0^1 P_t(j)^{1-\xi}\right]^{\frac{1}{1-\xi}}$.

- The downward sloping demand schedule is used by firm j to set $P_t(j)$ faced with $y_{D,t}(j)$ taking P_t and $Y_{D,t}$ as given.

Primitives: Staggered Price Setting of Output, II

- Calvo-staggered price setting firms face time-dependent updating to the optimal price $P_{C,t}$ at the probability $1-\mu_P$, or with probability μ_P, firms are stuck with P_{t-1} scaled by lagged inflation, ς_{t-1}.
  1. $\Longrightarrow$ Defines the aggregator or law of motion of the aggregate price level

$$P_t = \left[(1-\mu_P) P_{C,t}^{1-\xi} + \mu_P \left(\pi_{t-1} P_{t-1} \right)^{1-\xi} \right]^{\frac{1}{1-\xi}},$$

  2. where a firm unable to move to $P_{C,t}$ employs full indexation to lagged inflation, π_{t-1}, to adjust its price from date $t-1$ to date t.
  3. Calvo staggered price setting $\Longrightarrow$ nominal rigidities arise because different firms react differently to shocks at date t some of which cannot respond optimally to these shocks until these firms are able to alter their prices in the future.

- Since profit maximization is equivalent to cost minimization for firm j

$$\mathcal{TC}_t(j) = R_{K,t} K_t(j) + W_t N_t(j).$$

  where this firm's total nominal cost $\mathcal{TC}_t(j)$ is the sum of cost of factor inputs evaluated at the nominal rental rate of capital, $R_{K,t}$, and the nominal wage, W_t.

- Since final goods firms rent $K_t(j)$ in a perfectly competitive spot markets,
  1. $R_{K,t} = \Phi_t \psi y_t(j) / K_t(j)$, where Φ_t is the marginal cost of firm j during date t.
  2. $\Longrightarrow \Phi_t$ is the Lagrange multiplier attached to the production technology of firm j, which is common to all final goods firms.

Primitives: Staggered Price Setting of Output, III

- Although the labor market is monopolistic, the marginal product of labor is still $W_t = \Phi_t(1-\psi)y_t(j) \big/ [N_t(j) - N_0]$.
- Exploit the CRS technology common to all firms and the FONCs of firm j to find its total cost function $\mathcal{TC}_t(j) = \Phi_t y_{D,j,t} + W_t N_0 \Longrightarrow$ the net profit function (in real terms) of firm j

$$\mathcal{D}_{F,t}(j) = \left(\frac{P_t(j)}{P_t} - \phi_t\right)\left(\frac{P_t(j)}{P_t}\right)^{-\xi} Y_{D,t} - \left(\frac{W_t}{P_t}\right) N_0, \quad \phi_t \equiv \frac{\Phi_t}{P_t}. \qquad (\textbf{NK}.1)$$

- The consistency of the final goods price aggregator of P_t and its updating equation motivates imposing symmetry on final goods firms, $y_{D,t}(i) = y_{D,t}(j)$ for all i, j firms
  1. $\Longrightarrow$ identify $P_{C,t} = P_t(j)$ for all final goods firms that are able to change and commit to a new price at date t.
  2. As a result the problem of maximizing the expected discounted value of any firm's net profit function (**NK**.1) becomes

$$\mathbf{E}_t\left\{\sum_{i=0}^{\infty}(\beta\mu_P)^i\,\lambda_{t+i}\left[\left(\frac{P_{C,t}}{P_{t+i}} - \phi_{t+i}\right)\left(\frac{P_{C,t}}{P_{t+i}}\right)^{-\xi} Y_{D,t+i} - \left(\frac{W_{t+i}}{P_{t+i}}\right) N_0\right]\right\},$$

  3. where λ_t denotes the firm's time-varying discount factor (*i.e.*, the marginal utility of consumption because households own the final goods firms).

Primitives: Staggered Price Setting of Output, IV

- The problem of maximizing the firm's discounted profit function yields the optimal forward-looking inflation rate

$$\frac{P_{c,t}}{P_{t-1}} = \left(\frac{\xi}{\xi - 1}\right) \frac{\mathbf{E}_t \sum_{i=0}^{\infty} \left(\beta \mu_P\right)^i \lambda_{t+i} \phi_{t+i} Y_{D,t+i}\, \varsigma_{t+i}^{\xi}}{\mathbf{E}_t \sum_{i=0}^{\infty} \left(\beta \mu_P\right)^i \lambda_{t+i} Y_{D,t+i}\, \varsigma_{t+i}^{\xi-1}}, \qquad (\mathbf{NK}.2)$$

1. Inflation is smoothed by equation (**NK**.2) because an infinite cost of adjustment faces any firm unable to update its price to $P_{c,t}$ $\Longrightarrow$ there is no amount these firms can pay to optimally reset prices.
2. $\Longrightarrow$ Since prices or inflation do not adjust, shocks are propagated onto real variables generating the potential for a monetary transmission mechanism.

Primitives: Staggered Price Setting of Output, V

- Yun (1996) ties aggregate demand to aggregate supply, $Y_{A,t}$, to the supply price aggregator, $P_{A,t}^{-\xi} \equiv \int_0^1 P_t(j)^{-\xi} dj$, where its law of motion is

$$P_{A,t}^{-\xi} = (1-\mu)P_{C,t}^{-\xi} + \mu(\pi_{t-1}P_{A,t-1})^{-\xi}.$$

1. Aggregate output, $Y_{A,t} \equiv \int_0^1 y_{A,t}(j)dj$, and the downward sloping demand schedule for $y_{D,t}(j)$ set $Y_{D,t} = (P_t/P_{A,t})^{-\xi} Y_{A,t}$, which is a useful trick.
2. The trick is to eliminate $P_{C,t}$ from the state vector of the economy.
3. $\Longrightarrow$ Use the law of motion of $P_{A,t}$ to substitute for $P_{C,t}$ in the law of motion or aggregator of P_t, which will leave only P_t and P_{t-1} in the state vector.

- These facts yield aggregate real dividends

$$\frac{\mathcal{D}_{F,t}}{P_t} = \left(1-\psi\phi_t\right)\left(\frac{P_t}{P_{A,t}}\right)^{-\xi} Y_{A,t} - \left(\frac{R_{K,t}}{P_t}\right)K_t - \left(\frac{W_t}{P_t}\right)N_t.$$

1. The aggregate production function $Y_{A,t} = [u_tK_t]^{\psi}\left[(N_t-N_0)A_t\right]^{1-\psi}$ follows from the CRS technology, the relative prices $R_{K,t}/P_t$ and W_t/P_t, and
2. the definitions of the factor inputs $\Longrightarrow$ aggregate efficiency units of capital and labor services.

Primitives: Household Preferences and Its Budget Constraint

- There is a unit mass of households taking addresses on the unit circle, $\ell \in [0, 1]$.
- Household preferences are intertemporally separable and separable across (net) consumption, labor disutility, and real balances

$$\mathcal{U}\left(c_t, c_{t-1}, n_t(\ell), \frac{H_t}{P_t}\right) = \ln\left[c_t - hc_{t-1}\right] - \frac{\gamma}{1+\gamma} n_t(\ell)^{1+\frac{1}{\gamma}} + \ln\left[\frac{H_t}{P_t}\right], \quad (\mathbf{NK}.3)$$

1. where c_t = household consumption, $n_t(\ell)$ = the ℓth household's labor supply, H_t = household cash at the end of date $t-1$, and P_t = the aggregate price level,
2. the internal consumption habit parameter $h \in (0, 1)$, and the Frisch labor supply elasticity is $\gamma > 0$.

- The budget constraint of household ℓ is

$$\frac{H_{t+1}}{P_t} + \frac{B_{t+1}}{P_t} + c_t + x_t + a(u_t)k_t + \tau_t = r_t u_t k_t + \frac{W_t(\ell)}{P_t} n_t(\ell) + \frac{H_t}{P_t} + R_t \frac{B_t}{P_t} + \frac{\mathcal{D}_t}{P_t}, \quad (\mathbf{NK}.4)$$

1. where B_{t+1} = stock of government bonds carried from date t into date $t+1$, x_t = investment, k_{t+1} = household capital at the end of date t, τ_t = lump-sum government transfers, $r_t = R_{K,t}/P_t$, which is the real rental rate of k_t, $W_t(\ell)$ = the nominal wage of household ℓ, R_t = nominal return on B_t, and $a(u_t)$ is the capital utilization cost function, $a(1) = 0$, $a'(1) > 0$, and $a''(1) > 0$.
2. Change $u_t \Longrightarrow$ household ℓ has to forgo $a(\cdot)$ units of c_t per unit of k_t.

Primitives: The Law of Motion of Capital

- The law of motion of capital is

$$k_{t+1} = (1-\delta)k_t + \left[1 - \mathcal{S}\left(\frac{1}{\alpha}\frac{x_t}{x_{t-1}}\right)\right]x_t, \quad \delta \in (0, 1), \quad 0 < \alpha, \qquad (\textbf{NK}.5)$$

  1. where $\mathcal{S}(\cdot)$ denotes the cost of investment adjustment function, δ is the capital depreciation rate, α is deterministic TFP growth, and
  2. the cost function $\mathcal{S}(\cdot)$ is strictly convex, $\mathcal{S}(1) = \mathcal{S}'(1) = 0$ and $\mathcal{S}''(1) \equiv \varpi > 0$.

- The price of capital, q_t, is equivalent to the ratio of the replacement cost of capital to its market value, which is Tobin's q.
  1. Tobin's q exists in the NKDSGE model when the real friction of costly adjustment of investment binds.
  2. $\Longrightarrow q_t > 1$, when $\mathcal{S}(\cdot) > 0$ and $q_t = 1$, given $\mathcal{S}(\cdot) = 0$.

Primitives: Staggered Nominal Wage Setting

- The ℓth households charge firms $W_t(\ell)$ per unit of differentiated labor services in a monopolistic market in which a Calvo-staggered nominal wage mechanism operates.

  1. A wage elasticity $\theta > 0 \Longrightarrow$ aggregate labor supply $N_t = \left[\int_0^1 n_t(\ell)^{\frac{\theta-1}{\theta}} d\ell\right]^{\frac{\theta}{\theta-1}}$.
  2. Labor market monopoly $\Longrightarrow$ firms have downward-sloping labor demand schedules, $n_t(\ell) = \left[\frac{W_t}{W_t(\ell)}\right]^{\theta} N_t$, where $W_t = \left[\int_0^1 W_t(\ell)^{1-\theta} d\ell\right]^{\frac{1}{1-\theta}}$ is the nominal wage index and $W_t = \left[(1-\mu_W)W_{c,t}^{1-\theta} + \mu_W(\alpha^* \pi_{t-1} W_{t-1})^{1-\theta}\right]^{\frac{1}{1-\theta}}$ its aggregator.

- Under the Calvo-staggered nominal wage setting, changes in $W_t(\ell)$ are time-dependent.

  1. Households optimally update their $W_t(\ell)$s to $W_{c,t}$ at he probability $1-\mu_W$.
  2. Otherwise at probability μ_W, a household's $W_t(\ell)$ equals W_{t-1} indexed by steady-state TFP growth, $\alpha^* = \exp\{\alpha\}$, multiplied by the inflation rate, ς_{t-1} $\Longrightarrow$ full nominal wage indexation adjusted for balance growth.
  3. Calvo staggered nominal wage dynamics made operational in a NKDSGE model by Erceg, Henderson, and Levin (2000, "Optimal monetary policy with staggered wage and price contracts," *Journal of Monetary Economics* 46, 281–313).

A Household's Optimization Problem

- The ℓth household's dynamic optimization problem is

$$\textbf{Max}_{\{c_t,\, k_{t+1},\, H_{t+1},\, B_{t+1},\, W_t(\ell)\}} \mathbf{E}_t \left\{ \sum_{i=0}^{\infty} \beta^i \mathcal{U}\left(c_{t+i},\, c_{t+i-1},\, n_{t+i}(\ell),\, \frac{H_t}{P_t} \right) \right\}, \quad (\textbf{NK}.6)$$

  1. s.t. period utility (**NK**.3), the budget constraint (**NK**.4), the law of motion of capital (**NK**.5), and the downward-sloping labor demand, $n_t(\ell) = \left[\frac{W_t}{W_t(\ell)} \right]^{\theta} N_t$,
  2. given initial conditions $c_{-1},\ k_0,\ H_0$, and B_0.

- This problem yields the optimal nominal wage condition of the ℓth household

$$\left[\frac{W_{c,t}}{P_{t-1}} \right]^{1+\frac{\theta}{\gamma}} = \left(\frac{\theta}{\theta - 1} \right) \frac{\mathbf{E}_t \sum_{i=0}^{\infty} \left[\beta \mu_W \alpha^{*-\theta\left(1+\frac{1}{\gamma}\right)} \right]^i \left[\left[\frac{W_{t+i}}{P_{t+i-1}} \right]^{\theta} N_{t+i} \right]^{1+\frac{1}{\gamma}}}{\mathbf{E}_t \sum_{i=0}^{\infty} \left[\beta \mu_W \alpha^{*(1-\theta)} \right]^i \lambda_{t+i} \left[\frac{W_{t+i}}{P_{t+i-1}} \right]^{\theta} \left[\frac{P_{t+i}}{P_{t+i-1}} \right]^{-1} N_{t+i}}, \quad (\textbf{NK}.7)$$

  1. where λ_t is the Lagrange multiplier on the budget constraint (**NK**.4) $\Longrightarrow$ marginal utility of consumption.
  2. Equation (**NK**.7) smooths nominal wage growth, $\Longrightarrow$ forces labor supply to absorb TFP, monetary policy, and any other relevant shocks.
  3. Shifts in labor supply alter production and intra- and intertemporal margins $\Longrightarrow$ potential for endogenous propagation of real and nominal shocks.

The Government

- The NKDSGE model is closed using the government budget constraint

$$(1 + R_t)B_t + P_t\tau_t = [M_{t+1} - M_t] + B_{t+1}.$$

1. The government finances B_t, interest on B_t, and a lump-sum transfer τ_t (or if negative a lump-sum tax) with new bonds B_{t+1}, and money creation, $M_{t+1} - M_t$.
2. Assume government debt is in zero net supply, $B_{t+1} = 0$ and $P_t\tau_t = M_{t+1} - M_t$, along the equilibrium path at all dates t.

- Monetary policy is operated under a money growth or interest rate rule.
1. CEE (2005) identify monetary policy with a MA(∞) money growth process, which is equivalent to the AR(1) money growth rule

$$m_{t+1} = (1 - \rho_m)m^* + \rho_m m_t + \mu_t, \quad \mu_t \sim \mathcal{N}\left(0, \sigma_\mu^2\right), \qquad \textbf{(NK.8)}$$

2. where $m_{t+1} = \ln(M_{t+1}/M_t)$, m^* is its population mean, and $\left|\rho_m\right| < 1$.
3. The interest rate or Taylor rule reacts to changes in one-step ahead expected inflation, $\mathbf{E}_t\varsigma_{t+1}$, transitory output, $\tilde{Y}_t$, and the lagged policy rate, R_{t-1},

$$(1-\rho_R\mathbf{L})R_t = (1-\rho_R)\left(R^* + \kappa_\pi\mathbf{E}_t\pi_{t+1} + \kappa_y\tilde{Y}_t\right) + \upsilon_t, \quad \upsilon_t \sim \mathcal{N}\left(0, \sigma_\upsilon^2\right), \qquad \textbf{(NK.9)}$$

4. where $R^* = \exp(m^*)/\beta$ and $\left|\rho_R\right| < 1$, the Taylor principle, $1 < \kappa_\pi$ holds, $0 < \kappa_{\tilde{Y}}$, and $\mathbf{E}_t\pi_{t+1}$ and $\tilde{Y}_t$ are free of measurement error.
5. Non-systemic movements in monetary policy are generated by the money growth shock, μ_t, or the innovation to the policy rate, υ_t.

Equilibrium Definition for the NKDSGE Model

- The NKDSGE model is a decentralized economy in which equilibrium requires goods, labor, and money markets to clear in equilibrium.
  1. Goods market clears: $K_t = \int_0^1 K_t(j)dj = \int_0^1 k_t(\ell)d\ell$, given $0 < r_t$.
  2. Labor market clears: $N_t = \int_0^1 n_t(\ell)d\ell$, given $0 < W_t$.
  3. Money market clears: $M_t = H_t$, given $0 < P_t, R_t$.
  4. $\Longrightarrow$ The aggregate resource constraint is $Y_{A,t} = C_t + X_t + a(u_t)K_t$, where $C_t = c_t$ and $X_t = x_t$.
- Enforce a symmetric equilibrium on the markets in which final good firms and households have monopolistic power.
  1. Along the symmetric equilibrium path, firms i and j choose the same commitment price $P_{c,t} = P_t(i) = P_t(j) \Longrightarrow P_t = P_{A,t}$ and $Y_t = Y_{A,t}$.
  2. Similar restrictions on the nominal wages of households ℓ and $\imath$, $\Longrightarrow W_{c,t} = W_t(\ell) = W_t(\imath)$, where $W_{D,t}^{-\theta} = (1-\mu_W)W_{c,t}^{-\theta} + \mu_W\left(\alpha^*\pi_{t-1}W_{D,t-1}\right)^{-\theta}$ and $W_{D,t}^{-\theta} = \int_0^1 W_t(\ell)^{-\theta}d\ell$ deletes $W_{c,t}$ from the state and obtain $W_t = W_{D,t}$ and $N_t = n_t$.
- A rational expectations equilibrium (REE) equates, on average, firm and household subjective beliefs of r_t, and A_t with objective outcomes created by a market economy.
  1. The list includes μ_t and R_t under the money growth rule, (**NK**.8), or υ_t and R_t for the Taylor rule (**NK**.9).
  2. A flexible price regime adds P_t while a spot labor market appends W_t.

Optimality Conditions: The Household

- The NKDSGE models have FONC that are restricted by the primitives of preferences, technology, market structure, and monetary policy regime.
- The FONCs imply optimality and equilibrium conditions that must be satisfied by any candidate equilibrium time series.
- The optimality condition of consumption, which is the marginal utility of consumption is

$$\lambda_t = \frac{1}{C_t - hC_{t-1}} - \beta h \mathbf{E}_t \left\{ \frac{1}{C_{t+1} - hC_t} \right\}. \quad (\mathbf{OE}.1)$$

- The Euler equations for cash and the government bond are

$$\frac{\lambda_t}{P_t} = \beta \mathbf{E}_t \left\{ \frac{\lambda_{t+1}}{P_{t+1}} + \frac{1}{M_{t+1}} \right\}, \quad (\mathbf{OE}.2)$$

and

$$\frac{\lambda_t}{P_t} = \beta \mathbf{E}_t \frac{\lambda_{t+1}}{P_{t+1}} R_{t+1}. \quad (\mathbf{OE}.3)$$

Optimality Conditions: The Demand for Investment, Capital, and Labor

- The household's choice of investment leads to the optimality condition of investment

$$\frac{1-q_t}{q_t} + S\left(\frac{X_t}{\alpha^* X_{t-1}}\right) + S'\left(\frac{X_t}{\alpha^* X_{t-1}}\right)\frac{X_t}{\alpha^* X_{t-1}} = \frac{\beta}{\alpha^*}\mathbf{E}_t\left\{\frac{\lambda_{t+1}q_{t+1}}{\lambda_t q_t} S'\left(\frac{X_{t+1}}{\alpha^* X_t}\right)\left[\frac{X_{t+1}}{X_t}\right]^2\right\}. \qquad \textbf{(OE.4)}$$

- The Euler equation for capital is

$$q_t = \beta\mathbf{E}_t\left\{\frac{\lambda_{t+1}}{\lambda_t}\left[\psi u_{t+1}\phi_{t+1}\frac{Y_{A,t+1}}{K_{t+1}} - a(u_{t+1}) + q_{t+1}(1-\delta)\right]\right\}. \qquad \textbf{(OE.5)}$$

- Equating the cost of capital utilization with the firm's marginal product of capital implies a demand for efficiency unit of capital,

$$a'(u_t) = \psi\phi_t\frac{Y_{A,t}}{K_t}. \qquad \textbf{(OE.6)}$$

- The firm's intratemporal optimality condition implying labor demand is

$$\frac{W_t}{P_t} = \phi_t(1-\psi)\frac{Y_{A,t}}{N_t - N_0}. \qquad \textbf{(OE.7)}$$

Optimality and Equilibrium Conditions: Goods Market Pricing

- The solution to a monopolistically competitive firm's optimal pricing problem is

$$\frac{P_{c,t}}{P_{t-1}} = \left(\frac{\xi}{\xi-1}\right)\frac{\mathbf{E}_t\sum_{i=0}^{\infty}\left[\beta\mu_P\right]^i\lambda_{t+i}\phi_{t+i}Y_{D,t+i}\left[\frac{P_{t+i}}{P_{t+i-1}}\right]^{\xi}}{\mathbf{E}_t\sum_{i=0}^{\infty}\left[\beta\mu_P\right]^i\lambda_{t+i}Y_{D,t+i}\left[\frac{P_{t+i}}{P_{t+i-1}}\right]^{\xi-1}}. \qquad \textbf{(OE.8)}$$

- Laws of motion of the aggregate price level and aggregate supply price level are

$$P_t^{1-\xi} = \mu_P\left[\frac{P_{t-1}}{P_{t-2}}P_{t-1}\right]^{1-\xi} + (1-\mu_P)P_{c,t}^{1-\xi}, \qquad \textbf{(OE.9)}$$

and

$$P_{A,t}^{-\xi} = \mu_P\left[\frac{P_{A,t-1}}{P_{A,t-2}}P_{A,t-1}\right]^{-\xi} + (1-\mu_P)P_{c,t}^{-\xi}. \qquad \textbf{(OE.10)}$$

- A symmetric equilibrium in the goods market leads to

$$\frac{Y_{A,t}}{Y_{D,t}} = \left(\frac{P_{A,t}}{P_t}\right)^{-\xi}. \qquad \textbf{(OE.11)}$$

Optimality and Equilibrium Conditions: Nominal Wages

- A household sets its optimal nominal wage according to

$$\left[\frac{W_{c,t}}{P_{t-1}}\right]^{1+\frac{\theta}{\gamma}} = \left(\frac{\theta}{\theta-1}\right)\frac{\mathbf{E}_t\sum_{i=0}^{\infty}\left[\beta\mu_W\alpha^{*-\theta\left(1+\frac{1}{\gamma}\right)}\right]^i\left[\left[\frac{W_{t+i}}{P_{t+i-1}}\right]^{\theta}N_{t+i}\right]^{1+\frac{1}{\gamma}}}{\mathbf{E}_t\sum_{i=0}^{\infty}\left[\beta\mu_W\alpha^{*(1-\theta)}\right]^i\lambda_{t+i}\left[\frac{W_{t+i}}{P_{t+i-1}}\right]^{\theta}\left[\frac{P_{t+i}}{P_{t+i-1}}\right]^{-1}N_{t+i}}. \quad (\mathbf{OE}.12)$$

- The aggregate and aggregate demand nominal wages have laws of motion

$$W_t^{1-\theta} = \mu_W\left(\alpha^*\frac{P_{t-1}}{P_{t-2}}W_{t-1}\right)^{1-\theta} + (1-\mu_W)W_{c,t}^{1-\theta}, \quad (\mathbf{OE}.13)$$

and

$$W_{D,t}^{-\theta} = \mu_W\left(\alpha^*\frac{P_{t-1}}{P_{t-2}}W_{D,t-1}\right)^{-\theta} + (1-\mu_W)W_{c,t}^{-\theta}. \quad (\mathbf{OE}.14)$$

- The ratio of labor market demand and supply equal a ratio of the aggregate demand nominal wage to the aggregate nominal wage in a symmetric equilibrium

$$\frac{N_t}{n_t} = \left(\frac{W_{D,t}}{W_t}\right)^{-\theta}. \quad (\mathbf{OE}.15)$$

Miscellaneous Equilibrium Conditions

- Repeating the aggregate resource constraint

$$Y_{D,t} = C_t + X_t + a(u_t)K_t, \quad \textbf{(OE.16)}$$

1. the law of motion of capital

$$K_{t+1} = (1-\delta)K_t + \left[1 - \mathcal{S}\left(\frac{X_t}{\alpha^* X_{t-1}}\right)\right]X_t, \quad \textbf{(OE.17)}$$

2. and the aggregate supply or production function

$$Y_{A,t} = \Big[u_t K_t\Big]^{\psi}\Big[(N_t - N_0)A_t\Big]^{1-\psi}, \quad \textbf{(OE.18)}$$

3. completes the optimality and equilibrium conditions of the NKDSGE model.

- The impulse system or shocks driving the NKDSGE model consist of the money growth rule (**NK**.8) or the Taylor rule (**NK**.9) and the TFP process

$$\ln A_t = \alpha + \ln A_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2). \quad \textbf{(OE.19)}$$

Define Stochastic Detrending

- The TFP shock A_t creates trend movements or fluctuations in the level of the NKDSGE model.
  1. Aggregate quantities and prices are nonstationary because $\ln A_t$ is the random walk (with drift) process (**OE**.19).
  2. The state and endogenous flow variables need stochastic detrending to render stationary the equilibrium path of the NKDSGE model.
  3. Stochastic detrending consists of $\hat{Y}_{j,t} \equiv Y_{j,t}/A_t$, $j = A, D$, $\hat{C}_t \equiv C_t/A_t$, $\hat{X}_t \equiv X_t/A_t$, $\hat{K}_{t+1} \equiv K_{t+1}/A_t$, $\hat{P}_t \equiv P_t A_t/M_t$, $\hat{P}_{i,t} \equiv P_{i,t} A_t/M_t$, $i = A, c$, $\widehat{W}_t \equiv W_t/M_t$, and $\widehat{W}_{c,t} \equiv W_{\wp,t}/M_t$, $\wp = D, c$, and $\hat{\lambda}_t \equiv \lambda_t A_t$.
- Note nominal prices and wages are detrended by a nominal quantity, the money stock M_t, but relative prices, which includes real rates of return, are not.

The Results of Stochastic Detrending: The Household

- Apply these definitions to equations (**OE**.1)–(**OE**.14) to produce the stochastically detrended optimality and equilibrium conditions.

- The stochastically detrended marginal utility of consumption is

$$\widehat{\lambda}_t = \frac{\alpha_t}{\alpha_t \widehat{C}_t - h\widehat{C}_{t-1}} - \beta h \mathbf{E}_t \left\{ \frac{1}{\alpha_{t+1}\widehat{C}_{t+1} - h\widehat{C}_t} \right\}, \text{ where } \alpha_t = \frac{A_t}{A_{t-1}}. \qquad (\textbf{SD}.1)$$

- The money and bond Euler equations are

$$\frac{\widehat{\lambda}_t}{\widehat{P}_t} = \beta \mathbf{E}_t \left\{ \left[\frac{\widehat{\lambda}_{t+1}}{\widehat{P}_{t+1}} + 1 \right] \exp(-m_{t+1}) \right\}, \qquad (\textbf{SD}.2)$$

and

$$\frac{\widehat{\lambda}_t}{\widehat{P}_t} = \beta \mathbf{E}_t \left\{ \frac{\widehat{\lambda}_{t+1}}{\widehat{P}_{t+1}} \frac{R_{t+1}}{\exp(m_{t+1})} \right\}, \qquad (\textbf{SD}.3)$$

subsequent to stochastic detrending.

Stochastic Detrending: The Demand for Investment, Capital, and Labor

- Stochastic detrending of the intertemporal optimality conditions for investment and capital give

$$\frac{1-q_t}{q_t} + S\left(\frac{\alpha_t \widehat{X}_t}{\alpha^* \widehat{X}_{t-1}}\right) + S'\left(\frac{\alpha_t \widehat{X}_t}{\alpha^* \widehat{X}_{t-1}}\right)\frac{\alpha_t \widehat{X}_t}{\alpha^* \widehat{X}_{t-1}}$$

$$= \frac{\beta}{\alpha^*}\mathbf{E}_t\left\{\alpha_{t+1}\frac{q_{t+1}\widehat{\lambda}_{t+1}}{q_t\widehat{\lambda}_t} S'\left(\frac{\alpha_{t+1}\widehat{X}_{t+1}}{\alpha^* \widehat{X}_t}\right)\left[\frac{\widehat{X}_{t+1}}{\widehat{X}_t}\right]^2\right\}, \qquad (\mathbf{SD}.4)$$

and

$$q_t = \beta\mathbf{E}_t\left\{\frac{\widehat{\lambda}_{t+1}}{\widehat{\lambda}_t}\left[\psi u_{t+1}\phi_{t+1}\frac{\widehat{Y}_{t+1}}{\widehat{K}_{t+1}} + \frac{q_{t+1}[1-\delta] - a(u_{t+1})}{\alpha_{t+1}}\right]\right\}. \qquad (\mathbf{SD}.5)$$

- When applied to the intratemporal optimality conditions of capacity utilization and labor demand, the results are

$$a'(u_t) = \psi\phi_t\alpha_t\frac{\widehat{Y}_t}{\widehat{K}_t}, \qquad (\mathbf{SD}.6)$$

and

$$\frac{\widehat{W}_t}{\widehat{P}_t} = (1-\psi)\phi_t\frac{\widehat{Y}_t}{N_t - N_0}. \qquad (\mathbf{SD}.7)$$

Stochastic Detrending: Goods Market Pricing

- Stochastic detrending introduces money growth into the forward-looking optimally price of a final goods firm

$$\exp(m_t - \varepsilon_t)\frac{\widehat{P}_{c,t}}{\widehat{P}_{t-1}} = \left(\frac{\xi}{\xi-1}\right)\frac{\mathbf{E}_t\sum_{i=0}^{\infty}(\beta\mu_P)^i\widehat{\lambda}_{t+i}\phi_{t+i}\widehat{Y}_{D,t+i}\left[\exp(m_{t+i}-\varepsilon_{t+i})\frac{\widehat{P}_{t+i}}{\widehat{P}_{t+i-1}}\right]^{\xi}}{\mathbf{E}_t\sum_{i=0}^{\infty}(\beta\mu_P)^i\widehat{\lambda}_{t+i}\widehat{Y}_{D,t+i}\left[\exp(m_{t+i}-\varepsilon_{t+i})\frac{\widehat{P}_{t+i}}{\widehat{P}_{t+i-1}}\right]^{\xi-1}}. \quad (\mathbf{SD}.8)$$

- The demand schedule for final goods is little affected by stochastic detrending

$$\frac{\widehat{Y}_{A,t}}{\widehat{Y}_{D,t}} = \left(\frac{\widehat{P}_{A,t}}{\widehat{P}_t}\right)^{-\xi}. \quad (\mathbf{SD}.9)$$

- However, money growth appears in the laws of motion of the aggregate price level and the aggregate supply price because of stochastic detrending

$$\widehat{P}_t^{1-\xi} = \mu_P\left[\exp(-m_t + m_{t-1} + \varepsilon_t - \varepsilon_{t-1})\frac{\widehat{P}_{t-1}}{\widehat{P}_{t-2}}\widehat{P}_{t-1}\right]^{1-\xi} + (1-\mu_P)\widehat{P}_{c,t}^{1-\xi}, \quad (\mathbf{SD}.10)$$

and

$$\widehat{P}_{A,t}^{-\xi} = \mu_P\left[\exp(-m_t + m_{t-1} + \varepsilon_t - \varepsilon_{t-1})\frac{\widehat{P}_{A,t-1}}{\widehat{P}_{A,t-2}}\widehat{P}_{A,t-1}\right]^{-\xi} + (1-\mu_P)\widehat{P}_{c,t}^{-\xi}. \quad (\mathbf{SD}.11)$$

Stochastic Detrending: Nominal Wages, I

- After stochastic detrending, the optimal nominal wage is

$$\left[\exp(m_t)\frac{\widehat{W}_{c,t}}{\widehat{P}_{t-1}}\right]^{1+\frac{\theta}{\gamma}} = \left(\frac{\theta}{\theta-1}\right)$$

$$\times \frac{\mathbf{E}_t \sum_{i=0}^{\infty}\left(\beta\mu_W\right)^i \exp\left(\theta\left(1+\frac{1}{\gamma}\right)\right)\left(m_{t+i}+\sum_{j=1}^{i}\varepsilon_{t+j-1}\right)\left[\left[\frac{\widehat{W}_{t+i}}{\widehat{P}_{t+i-1}}\right]^{\theta} N_{t+i}\right]^{1+\frac{1}{\gamma}}}{\mathbf{E}_t \sum_{i=0}^{\infty}\left(\beta\mu_W\right)^i \lambda_{t+i}\exp\left(-(1-\theta)\left(m_{t+i}+\sum_{j=1}^{i}\varepsilon_{t+j-1}\right)\right)\left[\frac{\widehat{W}_{t+i}}{\widehat{P}_{t+i-1}}\right]^{\theta}\left[\frac{\widehat{P}_{t+i}}{\widehat{P}_{t+i-1}}\right]^{-1} N_{t+i}}. \quad \text{(SD.12)}$$

- Note that at $i = 0$, $\sum_{j=1}^{i} \varepsilon_{t+j-1} \equiv 1$.

- Money growth is a factor that can drive the household's choice of the stationary optimal nominal wage, $\widehat{W}_{c,t}$.

Stochastic Detrending: Nominal Wages, II

- Stochast detrending alter the laws of motion of the aggregate nominal wage and aggregate demand nominal wage to

$$\widehat{W}_t^{1-\theta} = \mu_W \left[\exp(-m_t + m_{t-1} - \varepsilon_{t-1}) \frac{\widehat{P}_{t-1}}{\widehat{P}_{t-2}} \widehat{W}_{t-1} \right]^{1-\theta} + (1-\mu_W)\widehat{W}_{c,t}^{1-\theta}, \qquad (\textbf{SD}.13)$$

and

$$\widehat{W}_{D,t}^{-\theta} = \mu_W \left[\exp(-m_t + m_{t-1} - \varepsilon_{t-1}) \frac{\widehat{P}_{t-1}}{\widehat{P}_{t-2}} \widehat{W}_{D,t-1} \right]^{-\theta} + (1-\mu_W)\widehat{W}_{c,t}^{-\theta}. \qquad (\textbf{SD}.14)$$

- The symmetric equilibrium and stochastic detrending sets the aggregate labor demand-aggregate labor supply ratio to the aggregate demand nominal wage-aggregate nominal wage ratio

$$\frac{N_t}{n_t} = \left(\frac{\widehat{W}_{D,t}}{\widehat{W}_t} \right)^{-\xi}. \qquad (\textbf{SD}.15)$$

Stochastic Detrending: Miscellaneous Equilibrium Conditions

- The stochastically detrended aggregate resource, law of motion of capital, and production technology are

$$\hat{Y}_t = \hat{C}_t + \hat{X}_t + \frac{a(u_t)\hat{K}_t}{\alpha_t}, \qquad \text{(SD.16)}$$

$$\hat{K}_{t+1} = \frac{(1-\delta)\hat{K}_t}{\alpha_t} + \left[1 - S\left(\frac{\alpha_t \hat{X}_t}{\alpha^* \hat{X}_{t-1}}\right)\right]\hat{X}_t, \qquad \text{(SD.17)}$$

and

$$\hat{Y}_t = \left[u_t \frac{\hat{K}_t}{\alpha_t}\right]^{\psi}\left[N_t - N_0\right]^{1-\psi}. \qquad \text{(SD.18)}$$

- The steady state equilibrium and the first-order linear approximation of the NKDSGE model is grounded on equations (**SD**.1)–(**SD**.18).

The Deterministic Steady State and the Household

- A deterministic steady state equilibrium sets all shocks innovations that drive fluctuations in the NKDSGE model to zero for all dates t.
  1. $\Longrightarrow \varepsilon_t = 0$ and $\mu_t = 0$ or $\upsilon_t = 0$ and also remember the restrictions on $u^* = 1$, $a(1) = 0$, and $S(1) = S'(1) = 0$ in the steady state equilibrium; see CEE (2005).
  2. Denote deterministic steady state values with λ^*, C^*, Y^*, X^*, N^*, K^*, q^*, W^*, r^*, P^*, u^*, ϕ^*, and R^*, which correspond to the associated endogenous variables found in the stochastically detrended optimality and equilibrium conditions (**SD**.1)–(**SD**.18).

- The deterministic steady state conditions for the household's consumption function, money demand, and bond demand are

$$C^*\lambda^* = \frac{\alpha^* - \beta h}{\alpha^* - h}, \qquad \textbf{(STEDST.1)}$$

$$\frac{\lambda^*}{P^*} = \frac{\beta}{\exp(m^*) - \beta}, \qquad \textbf{(STEDST.2)}$$

and

$$R^* = \frac{\exp(m^*)}{\beta}. \qquad \textbf{(STEDST.3)}$$

The Deterministic Steady State and Investment, Capital, and Labor

- The demand for investment, capital, and labor become

$$\frac{K^*}{Y^*} = \frac{\beta\alpha^*\psi\phi^*}{\alpha^* - \beta(1-\delta)}, \tag{STEDST.4}$$

$$q^* = 1, \tag{STEDST.5}$$

$$a'(1) = \psi\phi^*\alpha^*\frac{Y^*}{K^*}, \tag{STEDST.6}$$

$$\frac{W^*}{P^*} = (1-\psi)\phi^*\frac{Y^*}{N^* - N_0}, \tag{STEDST.7}$$

in the deterministic steady state.

The Deterministic Steady State and Pricing in the Goods and Labor Markets

- In the deterministic steady state, pricing in the goods markets collapses to

$$\phi^* = \frac{\xi - 1}{\xi}, \qquad (\textbf{STEDST}.8)$$

which is steady state real marginal cost and because of the symmetric equilibrium

$$P^* = P_A^* = P_c^* \text{ and } Y^* = Y_A^*. \qquad (\textbf{STEDST}.9)$$

- Similarly, the steady state nominal wage is found from

$$W^* = \left(\frac{\theta}{\theta - 1}\right)\frac{P^*}{\lambda^*}N^{*\frac{1}{\gamma}}, \qquad (\textbf{STEDST}.10)$$

where the symmetric equilibrium gives

$$W^* = W_D^* = W_c^*. \qquad (\textbf{STEDST}.11)$$

The Deterministic Steady State: Miscellaneous Equilibrium Conditions

- The aggregate resource constraint, the law of motion of capital, and the production technology are restricted by the deterministic steady state restricts to

$$Y^* = C^* + X^*, \qquad \textbf{(STEDST.12)}$$

$$\frac{X^*}{K^*} = 1 - \frac{1-\delta}{\alpha^*}, \qquad \textbf{(STEDST.13)}$$

and

$$Y^* = \left(\frac{K^*}{\alpha^*}\right)^{\psi}\left[N^* - N_0\right]^{1-\psi}. \qquad \textbf{(STEDST.14)}$$

The Deterministic Steady State: Notes, I

- The deterministic steady state equilibrium is a flexible price monopolistically competitive equilibrium in which there are real adjustment costs.
  1. The real frictions are internal consumption habit afflicting households $(h = 0)$ $\Longrightarrow$ the steady state consumption function (**STEDST**.1),
  2. the fixed labor cost $(N_0 > 0)$ imposed on the production technology of the final goods firms $\Longrightarrow$ sustains monopolistic competition, and
  3. capacity utilization costs inflicted on households $\Longrightarrow$ marginal product of capital is restricted by the ratio of the marginal cost of adjusting capacity utilization, $a'(1) > 0$, deflated by steady state marginal cost, ϕ^*, defined by equation (**STEDST**.8).
  4. but this real cost is absent from the deterministic steady state aggregate resource constraint (**STEDST**.12) $\Longrightarrow a(1) = 0$.
  5. However, the cost of adjusting investment is zero in the deterministic steady state $\Longrightarrow S(1) = S'(1) = 0$.
- There are no nominal frictions in the deterministic steady state equilibrium.
- Thus, the NKDSGE model responds to a given shock with a transition path restricted by sticky final goods prices and nominal wages, but ends at a steady state equilibrium lacking nominal frictions.

The Deterministic Steady State: Notes, II

- The endogenous steady state ratios and variables are functions of the NKDSGE model's deep parameters, $\Xi = [\beta \;\; h \;\; \gamma \;\; \alpha \;\; \psi \;\; N_0 \;\; \delta \;\; \xi, \;\; \theta \;\; m^*]'$.
- Find N^* by applying nonlinear equation solver to steady state conditions (**STEDST**.7), (**STEDST**.10), and (**STEDST**.14) $\Longrightarrow$ given N^*, use equations (**STEDST**.10) and (**STEDST**.2) to compute W^*.
- Next, combine equations (**STEDST**.1) and (**STEDST**.2) to calculate P^*C^*, which together with equation (**STEDST**.7) produces C^*/Y^*, given N^* and W^*.
- Divide both sides of the equality of the steady state aggregate resource constraint (**STEDST**.12) by Y^* to solve for X^*/Y^*, given the C^*/Y^* ratio obtained in the previous step.
- Since $\frac{X^*}{Y^*} = \frac{K^*}{Y^*}\frac{X^*}{K^*}$, divide both sides of the equality of the steady state production function (**STEDST**.14) and combine the result with equation (**STEDST**.13) to find $\frac{K^*}{Y^*}$ $\Longrightarrow$ equation (**STEDST**.6) gives $a'(1)$.
- Given the $\frac{K^*}{Y^*}$ ratio, the steady state production function (**STEDST**.14) yields K^*, which in turn generates Y^* $\Longrightarrow$ calculate C^*, X^*, P^*, and λ^*.

The Deterministic Steady State: Notes, III

- Endogenous deterministic steady state ratios and levels grounded in population relationships.
    1. Calibrate Ξ to generate unconditional population first moments of the steady state ratios and levels.
    2. Given unconditional sample first moments of $\frac{C_t}{Y_t}$, $\frac{X_t}{Y_t}$, etc., estimate several of the deep structural parameters located in Ξ.
    3. There may be an insufficient number of sample moments to identify the ten elements of Ξ.

- How to choose which elements of Ξ to estimate? $\Longrightarrow$ Are some deep structural parameters more economically compelling than others?

Linearizing the NKDSGE Model: Introduction

- When likelihood estimation of NKDSGE models is grounded on the Kalman filter, require linear approximate decisions rule for the endogenous state variables of the NKDSGE model.

- Log linearize the optimality and equilibrium conditions of the NKDSGE model to create these linear approximate decision rules.
  1. Compute first-order Taylor expansions of the stochastically detrended system (**SD**.1)–(**SD**.18) around the deterministic steady state given by equations (**STEDST**.1)–(**STEDST**.14).
  2. Construction of the log linear approximations exploit, for example, the definitions $\tilde{C}_t = \ln \hat{C}_t - \ln C^*$ or $\tilde{N}_t = \ln N_t - \ln N^*$.
  3. Since $\dfrac{P_t}{P_{t-1}} = \dfrac{\tilde{P}_t \ M_t \ A_{t-1}}{\tilde{P}_{t-1} M_{t-1} A_t}$, steady state inflation, $\pi^* = m^* - \alpha$.

- Remember a symmetric equilibrium has several implications for the log linear approximation of the NKDSGE model.
  1. Subsequent to log linearizing around the deterministic steady state,
  2. the assumption of a symmetric equilibrium equates the aggregate price indexes $\tilde{P}_t = \tilde{P}_{A,t}$ and the aggregate nominal wages $\widetilde{W}_t = \widetilde{W}_{D,t}$, given the initial conditions $P_0 = P_{A,0}$ and $W_0 = W_{D,0}$.
  3. $\Longrightarrow$ Reduces the dimension of the state vector.

Linearizing the NKDSGE Model: The Household Optimality Conditions

- Apply log linearization to the stochastically detrended household Euler marginal utility, cash, and bond conditions (**SD**.1), (**SD**.2), and (**SD**.8) to obtain the second-order expectational difference equation in $\widetilde{C}_t$

$$(\alpha^* - h)(\alpha^* - \beta h)\widetilde{\lambda}_t = \beta\alpha^* h \mathbf{E}_t \widetilde{C}_{t+1} - (\beta h^2 + \alpha^{*2})\widetilde{C}_t + \alpha^* h(\widetilde{C}_{t-1} - \varepsilon_t), \quad (\mathbf{LZ}.1)$$

and the first-order expectational difference equations in $\widetilde{\lambda}_t$ and $\widetilde{P}_t$

$$\widetilde{\lambda}_t - \widetilde{P}_t = \frac{\lambda^*}{\lambda^* + P^*}\mathbf{E}_t\left\{\widetilde{\lambda}_{t+1} - \widetilde{P}_{t+1}\right\} - \widetilde{m}_{t+1}, \quad (\mathbf{LZ}.2)$$

and

$$\widetilde{\lambda}_t - \widetilde{P}_t = \mathbf{E}_t\left\{\widetilde{\lambda}_{t+1} - \widetilde{P}_{t+1} + \widetilde{R}_{t+1}\right\} - \widetilde{m}_{t+1}. \quad (\mathbf{LZ}.3)$$

- Remember m_{t+1} is realized at the end of date t.

Linearizing the NKDSGE Model: The Demand for Investment, Capital, & Labor

- The stochastically detrended optimality conditions (**SD**.4), (**SD**.5), (**SD**.6) and (**SD**.7) yield log linearized conditions, which are
  1. a second-order expectational difference equation in $\tilde{X}_t$

$$\beta\varpi\mathbf{E}_t\tilde{X}_{t+1} - (1+\beta)\varpi\tilde{X}_t + \varpi\tilde{X}_{t-1} + \tilde{q}_t = \varpi\varepsilon_t, \qquad (\mathbf{LZ}.4)$$

  2. a second-order expectational difference equation in $\tilde{q}_t$

$$\tilde{q}_t + \tilde{\lambda}_t = \mathbf{E}_t\left\{\tilde{\lambda}_{t+1} + \beta\psi\phi^*\frac{Y^*}{K^*}\left[\tilde{\phi}_{t+1} + \tilde{Y}_{t+1} - \tilde{K}_{t+1}\right] + \beta\frac{1-\delta}{\alpha^*}\tilde{q}_{t+1}\right\}, \qquad (\mathbf{LZ}.5)$$

  3. an intratemporal condition relating deviations in capacity utilization from its deterministic steady state value, $\tilde{u}_t$, to an equivalent notion of the real marginal cost of increasing the capital input by one unit plus the TFP shock innovation

$$\varrho\tilde{u}_t = \tilde{\phi}_t + \tilde{Y}_t - \tilde{K}_t + \varepsilon_t, \quad \varrho \equiv \frac{a''(1)}{a'(1)}, \qquad (\mathbf{LZ}.6)$$

  4. and an intratemporal condition relating deviations in the real wage from its deterministic steady state to the linearized marginal cost of adding an additional unit of labor input to production

$$\widetilde{W}_t - \tilde{P}_t = \tilde{\phi}_t + \tilde{Y}_t - \frac{N^*}{N^* - N_0}\tilde{N}_t. \qquad (\mathbf{LZ}.7)$$

Linearizing the NKDSGE Model: Sticky Prices and Nominal Wages

- Log linearization of the optimal price and nominal wage conditions (**SD**.8) and (**SD**.12) produces second-order expectational difference equations in $\tilde{\pi}_t$ and $\widetilde{W}_t$

$$\mu_P(1+\beta)\tilde{\pi}_t = \beta\mu_P \mathbf{E}_t\tilde{\pi}_{t+1} + \mu_P\tilde{\pi}_{t-1} + (1-\mu_P)(1-\beta\mu_P)\tilde{\phi}_t$$
$$+ \beta\mu_P\widetilde{m}_{t+1} - \mu_P(1+\beta)(\widetilde{m}_t - \varepsilon_t) + \mu_P(\widetilde{m}_{t-1} - \varepsilon_{t-1}), \qquad (\mathbf{LZ}.8)$$

and

$$\left[1 + \beta\mu_W^2 - \frac{\theta(1-\mu_W)(1-\beta\mu_W)}{\theta+\gamma}\right]\widetilde{W}_t = \beta\mu_W\mathbf{E}_t\widetilde{W}_{t+1} + \mu_W\widetilde{W}_{t-1}$$
$$+ \left[\frac{(1-\mu_W)(1-\beta\mu_W)}{\theta+\gamma}\right]\tilde{N}_t - \left[\frac{\gamma(1-\mu_W)(1-\beta\mu_W)}{\theta+\gamma}\right]\left(\tilde{\lambda}_t - \tilde{P}_t\right) - \beta\mu_W\tilde{\pi}_t + \mu_W\tilde{\pi}_{t-1}$$
$$+ \beta\mu_W\widetilde{m}_{t+1} - (1+\beta)\mu_W\widetilde{m}_t + \mu_W\widetilde{m}_{t-1} + \beta\mu_W\varepsilon_t - \mu_W\varepsilon_{t-1}. \qquad (\mathbf{LZ}.9)$$

- The bivariate system of inflation, $\tilde{\pi}_t \equiv \tilde{P}_t - \tilde{P}_{t-1}$, and the nominal wage is recursive $\Longrightarrow$ compute $\widetilde{W}_t$ conditional on $\tilde{\pi}_t$, given $\tilde{N}_t$ and shocks to TFP and money growth.
- The second-order expectational difference equation (**LZ**.9) has a Phillips curve-like relationship between $\tilde{\pi}_t$ and $\tilde{N}_t \Longrightarrow$ that is conditional on $\widetilde{W}_t$, its one-step ahead expectation, and lag, and shocks to money growth and TFP.

Linearizing the NKDSGE Model: Miscellaneous Equilibrium Conditions

- The stochastically detrended aggregate resource constraint, law of motion of capital, and aggregate technology, which are equations (**SD**.13), (**SD**.14), and (**SD**.15), are after log linearization

$$\widetilde{K}_{t+1} \;=\; \frac{1-\delta}{\alpha^*}\left(\widetilde{K}_t - \varepsilon_t\right) \;+\; \frac{X^*}{K^*}\widetilde{X}_t, \qquad (\mathbf{LZ}.10)$$

$$\widetilde{Y}_t \;=\; \frac{C^*}{Y^*}\widetilde{C}_t \;+\; \frac{X^*}{Y^*}\widetilde{X}_t \;+\; \psi\phi^*\widetilde{u}_t, \qquad (\mathbf{LZ}.11)$$

and

$$\widetilde{Y}_t \;=\; \psi\left(\widetilde{u}_t + \widetilde{K}_t\right) \;+\; (1-\psi)\frac{N^*}{N^*-N_0}\widetilde{N}_t \;-\; \psi\varepsilon_t, \qquad (\mathbf{LZ}.12)$$

Linearizing the NKDSGE Model: The Impulse System

- The TFP process (**OE**.19) is recast in stationary log linear form as

$$\ln \alpha_t - \alpha = \varepsilon_t. \qquad (\textbf{LZ}.13)$$

- The log linear money growth rule (**NK**.8) is written

$$\widetilde{m}_{t+1} = \rho_m \widetilde{m}_t + \mu_t. \qquad (\textbf{LZ}.14)$$

- When a log linearized NKDSGE model operates under a interest rate rule, the revised Taylor rule shows deviations of the policy rate from its deterministic steady state react to $\widetilde{m}_{t+1}$ besides $\mathbf{E}_t \widetilde{\pi}_{t+1}$ and $\widetilde{Y}_t$

$$(1-\rho_R \mathbf{L})\widetilde{R}_t = (1-\rho_R)\Big(\kappa_\pi \mathbf{E}_t \widetilde{\pi}_{t+1} + \kappa_\pi \widetilde{m}_{t+1} + \kappa_y \widetilde{Y}_t\Big) + \upsilon_t. \qquad (\textbf{LZ}.15)$$

Linearizing the NKDSGE Model: The Log Linearized System

- Collect the unknowns of the linear approximate NKDSGE model consisting of the linear stochastic difference equations (**LZ**.1)–(**LZ**.12), the TFP shock innovation (**LZ**.13), and the Taylor rule (**LZ**.15) in

$$\mathcal{S}_t = \left[\widetilde{\lambda}_t \;\; \widetilde{C}_t \;\; \widetilde{X}_t \;\; \widetilde{q}_t \;\; \widetilde{Y}_t \;\; \widetilde{R}_t \;\; \widetilde{u}_t \;\; \widetilde{\phi}_t \;\; \widetilde{N}_t \;\; \widetilde{m}_{t+1} \;\; \widetilde{K}_{t+1} \;\; \mathbf{E}_t\widetilde{P}_{t+1} \;\; \mathbf{E}_t\widetilde{W}_{t+1} \;\; \widetilde{P}_t \;\; \widetilde{W}_t\right]'.$$

- The endogenous and exogenous variables are collected in $\mathcal{S}_t \Longrightarrow$ all the stochastically detrended and demeaned NKDSGE models variables.

  1. The vector $\mathcal{S}_t$ includes the 'true' state variable $\widetilde{K}_{t+1}$ and the 'artificial' state variables $\mathbf{E}_t\widetilde{P}_{t+1}$, $\mathbf{E}_t\widetilde{W}_{t+1}$, $\widetilde{P}_t$, and $\widetilde{W}_t$.
  2. $\Longrightarrow$ Price and nominal wage expectations and the aggregate price level and nominal wage contribute to the state of the economy.

Linearizing the NKDSGE Model: A Note on the Money Market Equilibrium

- Under the Taylor rule (**LZ**.15), it and the linearized Euler equations for cash and bonds jointly restrict the money market equilibrium of this NKDSGE model.
  1. Equate the linearized Euler equations (**LZ**.2) and (**LZ**.3) for cash and bonds $\Longrightarrow \mathbf{E}_t\left\{\tilde{\lambda}_{t+1} - \tilde{P}_{t+1}\right\} = -\exp(m^*)/[\exp(m^*) - \beta]\mathbf{E}_t\tilde{R}_{t+1}$.
  2. The linearized bond Euler equation (**LZ**.3) is a Fisher-like equation, $\widetilde{m}_{t+1} = \mathbf{E}_t\left\{\tilde{R}_{t+1} - \tilde{\pi}_{t+1}\right\} - \mathbf{E}_t\left\{-(\tilde{\lambda}_{t+1} - \tilde{\lambda}_t)\right\}$, where money demand is a residual between the ' ex ante real rate' and $-(\tilde{\lambda}_{t+1} - \tilde{\lambda}_t)$, which represents the SDF.

- Under the money growth rule (**LZ**.14), the bond Euler equation is redundant $\Longrightarrow$ drop the linearized bond Euler equation (**LZ**.3) from the system and delete $\tilde{R}_t$ from $\mathcal{S}_t$.
  1. 'Money demand' equates the marginal utility of consumption-price level differential to the present discounted expected value of money growth using the linearized money Euler equation (**LZ**.2),

$$\tilde{\lambda}_t - \tilde{P}_t = -\sum_{j=0}^{\infty}\left[\frac{\beta}{\exp(m^*)}\right]^j \mathbf{E}_t\widetilde{m}_{t+1+j}.$$

  2. Compute $\mathbf{E}_t\tilde{R}_{t+1}$ as a 'residual' from the linearized bond Euler equation (**LZ**.3)

$$\mathbf{E}_t\tilde{R}_{t+1} = \mathbf{E}_t\left\{-\left(\tilde{\lambda}_{t+1} - \tilde{P}_{t+1}\right) + \tilde{\lambda}_t - \tilde{P}_t\right\} + \widetilde{m}_{t+1}.$$

Solving the Linearized NKDSGE Model: Introduction

- Solve the linear approximate NKDSGE model for $\mathcal{S}_t$, which consists of the linear stochastic difference equations (**LZ**.1)–(**LZ**.12) and the Taylor rule (**LZ**.15) using methods developed by
  1. Sims (2002, "Solving linear rational expectations models," *Computational Economics* 20, 1–20).
  2. Also see Zadrozny (1998, "An eigenvalue method of undetermined coefficients for solving linear rational expectations models," *Journal of Economic Dynamics and Control* 22, 1353–1373).
  3. Their approaches are explicit about handling expectations (or expectational errors) to solve linear RE models.

- Sims' solution algorithm needs the expectational forecast errors

$$\begin{array}{llllllllll} \vartheta_{\tilde{\lambda},t} & = & \tilde{\lambda}_t - E_{t-1}\tilde{\lambda}_t, & \vartheta_{\tilde{C},t} & = & \tilde{C}_t - E_{t-1}\tilde{C}_t, & \vartheta_{\tilde{X},t} & = & \tilde{X}_t - E_{t-1}\tilde{X}_t, \\ \vartheta_{\tilde{q},t} & = & \tilde{q}_t - E_{t-1}\tilde{q}_t, & \vartheta_{\tilde{Y},t} & = & \tilde{Y}_t - E_{t-1}\tilde{Y}_t, & \vartheta_{\tilde{u},t} & = & \tilde{u}_t - E_{t-1}\tilde{u}_t, \\ \vartheta_{\tilde{\phi},t} & = & \tilde{\phi}_t - E_{t-1}\tilde{\phi}_t, & \vartheta_{\tilde{N},t} & = & \tilde{N}_t - E_{t-1}\tilde{N}_t, & \vartheta_{\tilde{P},t} & = & \tilde{P}_t - E_{t-1}\tilde{P}_t, \end{array}$$

  and $\vartheta_{\widetilde{W},t} = \widetilde{W}_t - E_{t-1}\widetilde{W}_t$.

- Collect these forecast errors into the vector

$$\vartheta_t = \left[\vartheta_{\tilde{\lambda},t} \;\; \vartheta_{\tilde{C},t} \;\; \vartheta_{\tilde{X},t} \;\; \vartheta_{\tilde{q},t} \;\; \vartheta_{\tilde{Y},t} \;\; \vartheta_{\tilde{R},t} \;\; \vartheta_{\tilde{u},t} \;\; \vartheta_{\tilde{\phi},t} \;\; \vartheta_{\tilde{N},t} \;\; \vartheta_{\tilde{P},t} \;\; \vartheta_{\widetilde{W},t}\right]'.$$

Solving the Linearized NKDSGE Model: Steps 1 and 2

- Constructing the solution to the linear approximate optimality and equilibrium conditions (**LZ**.1)–(**LZ**.12) and the Taylor rule (**LZ**.15) is a multi-step process.

- Define $\zeta_t = [\varepsilon_t \;\; \upsilon_t]'$ and remember $\mathcal{S}_t$ contains elements that appear in equations (**LZ**.1)–(**LZ**.12) as one-step ahead expectations.
  1. When monetary policy is the AR(1) money growth rule (**LZ**.14), $\zeta_t = [\varepsilon_t \;\; \mu_t]'$.
  2. If there are additional shocks, say preference, mark-up, and government spending shocks, the innovations to these shocks would be added to ζ_t while the demeaned logs of these shocks would be added to $\mathcal{S}_t$.

- Next, construct the multivariate first-order stochastic difference equation system of the NKDSGE model

$$\mathbb{G}_0\mathcal{S}_t = \mathbb{G}_1\mathcal{S}_{t-1} + \mathbb{V}\zeta_t + \mathbb{K}\vartheta_t, \qquad (\textbf{SLZ}.1)$$

where $\mathbb{G}_0$, $\mathbb{G}_1$, $\mathbb{V}$, and $\mathbb{K}$ express the cross-equation restrictions embedded in the optimality and equilibrium conditions (**LZ**.1)–(**LZ**.12), and the Taylor rule (**LZ**.15).

1. Cross-equation restrictions are the hallmark of linear RE models.
2. Solving linearized NKDSGE models is difficult because S_t contains forward looking expectations and $\mathbb{G}_0$ is often has reduced rank (*i.e.*, singular).
3. $\Longrightarrow$ Cannot compute the eigenvalues of the equations (**SLZ**.1) from $\mathbb{G}_0^{-1}\mathbb{G}_1$.

Solving the Linearized NKDSGE Model: The QZ Decomposition and Step 3

- Sims (2002) studies multivariate linear RE models similar to (**SLZ**.1).
  1. The solution algorithm taps the QZ (or generalized complex Schur) decomposition of the square matrices $\mathbb{G}_0$ and $\mathbb{G}_1$.
  2. Assume the eigenvalues of $\mathbb{G}_0$ and $\mathbb{G}_1$ reside on the open interval $(0, \infty)$.
  3. $\Longrightarrow$ Compute the eigenvalues of the pair $\left[\mathbb{G}_0, \ \mathbb{G}_1\right]$, where $\mathbb{G}_0$ can be singular.
  4. The QZ decomposition employs $\mathbb{Q}'\mathbb{F}\mathbb{Z}' = \mathbb{G}_0$ and $\mathbb{Q}'\mathbb{O}\mathbb{Z}' = \mathbb{G}_1$, where $\mathbb{Q}'\mathbb{Q} = \mathbb{Z}'\mathbb{Z} = \mathbf{I}$, $\mathbb{F}$ and $\mathbb{O}$ are upper triangular matrices, and there are possibly complex elements in the square matrices $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{F}$ and $\mathbb{O}$.
  5. Define $\mathbb{D}_t = \mathbb{Z}'\mathcal{S}_t$ and let $\mathbb{Q}_{j\cdot}$ denote the jth block of rows of $\mathbb{Q}$.

- Premultiply the system of linear equations (**SLZ**.1) by $\mathbb{Q}$, remember $\mathbb{F}$ is a square upper triangular matrix, and break $\mathbb{D}_t$ into $\mathbb{D}_{1,t} = \mathbb{Z}'_{\cdot 1}\mathcal{S}_t$, $\mathbb{D}_{2,t} = \mathbb{Z}'_{\cdot 2}\mathcal{S}_t$, where $\mathbb{Z} = \left[\mathbb{Z}_{\cdot 1} \ \ \mathbb{Z}_{\cdot 2}\right]$, which maps the dynamic system $\mathbb{F}\mathbb{D}_t = \mathbb{O}\mathbb{D}_{t-1} + \mathbb{Q}\mathbb{V}\zeta_t + \mathbb{Q}\mathbb{K}\vartheta_t$ into

$$\begin{bmatrix} \mathbb{F}_{11} & \mathbb{F}_{12} \\ \mathbf{0} & \mathbb{F}_{22} \end{bmatrix}\begin{bmatrix} \mathbb{D}_{1,t} \\ \mathbb{D}_{2,t} \end{bmatrix} = \begin{bmatrix} \mathbb{O}_{11} & \mathbb{O}_{12} \\ \mathbf{0} & \mathbb{O}_{22} \end{bmatrix}\begin{bmatrix} \mathbb{D}_{1,t-1} \\ \mathbb{D}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \mathbb{Q}_{1\cdot} \\ \mathbb{Q}_{2\cdot} \end{bmatrix}\left(\mathbb{V}\zeta_t + \mathbb{K}\vartheta_t\right). \qquad (\mathbf{SLZ}.2)$$

- A QZ decomposition of $\mathbb{G}_0$ and $\mathbb{G}_1$ always exists, but these eigenvalues are not unique.
  1. The generalized eigenvalues of $\left[\mathbb{F} \ \ \mathbb{O}\right]$ can be unique, where these eigenvalues are denoted $f_{ii}^{-1}o_{ii}$ and f_{ii} and o_{ii} are diagonal elements of these matrices.
  2. The largest elements of $\mathbb{F}$ are placed in $\mathbb{F}_{22}$ $\Longrightarrow$ the lower right block of the system (**SLZ**.2).

Solving the Linearized NKDSGE Model: Step 4a

- A subset of the generalized eigenvalues $f_{ii}^{-1}o_{ii} \geq \overline{\nu}$ and the remaining $f_{ii}^{-1}o_{ii} < \overline{\nu}$, where $\overline{\nu} > 1$ is the maximal growth rate of the elements of $\mathcal{S}_t$.
  1. The eigenvalues $f_{ii}^{-1}o_{ii}$ are ordered to partition the system (**SLZ**.2) in such a way to place only explosive forward-looking elements in $\mathbb{D}_{2,t}$.
  2. $\Longrightarrow$ Exogenous shocks that are intrinsic and potentially extrinsic (*i.e.*, sunspots).

- The partition of the system of equations (**SLZ**.2) creates (what Sims calls) the 'reduced form' process of $\mathbb{D}_{2,t}$, which is the second row of the system (**SLZ**.2)

$$\mathbb{D}_{2,t} = \boldsymbol{\mathcal{M}}\mathbb{D}_{2,t-1} + \boldsymbol{\mathcal{M}}\mathbb{O}_{22}^{-1}\mathbb{Q}_{2.}(\mathbb{V}\zeta_t + \mathbb{K}\vartheta_t), \text{ where } \boldsymbol{\mathcal{M}} \equiv \mathbb{F}_{22}^{-1}\mathbb{O}_{22}. \quad (\textbf{SLZ}.3)$$

- Iterate equation (**SLZ**.3) forward to find

$$\mathbb{D}_{2,t} = -\sum_{i=0}^{\infty}\boldsymbol{\mathcal{M}}^{-i}\mathbb{O}_{22}^{-1}\mathbb{Q}_{2.}(\mathbb{V}\zeta_{t+i+1} + \mathbb{K}\vartheta_{t+i+1}), \quad (\textbf{SLZ}.4)$$

after invoking the transversality condition $\lim_{i\longrightarrow\infty}\boldsymbol{\mathcal{M}}^{-i}\mathbb{D}_{2,t+i} = 0$ holds.

Solving the Linearized NKDSGE Model: Step 4b

- The goal is to exclude extrinsic equilibria from the solution of the expected present discounted value formula (**SLZ**.4) of $\mathbb{D}_{2,t}$.
- Remember extrinsic shocks are subjective beliefs held by the households and firms of the linearized NKDSGE model about the behavior of the elements of ϑ_t.
- Subjective beliefs do not matter for the expected present discounted value (**SLZ**.4) if
  1. $\mathbb{D}_{1,t}$ and $\mathbb{D}_{2,t}$ are orthogonal to movements in the expectational error vector ϑ_t.
  2. Since there are no subjective beliefs producing persistent expectational errors, $\mathbb{D}_{2,t}$ belongs only to the date t information set.
- When only the intrinsic shocks of ζ_t drive fluctuations in the linearized NKDSGE model, the expected present discounted value formula (**SLZ**.4) is

$$\mathbf{E}_t \sum_{i=0}^{\infty} \mathcal{M}^{-i} \mathbb{O}_{22}^{-1} \mathbb{Q}_{2.} \mathbb{V} \zeta_{t+i+1} = \sum_{i=0}^{\infty} \mathcal{M}^{-i} \mathbb{O}_{22}^{-1} \mathbb{Q}_{2.} \left(\mathbb{V} \zeta_{t+i+1} + \mathbb{K} \vartheta_{t+i+1} \right), \qquad (\mathbf{SLZ}.5)$$

  because $\mathbf{E}_t \vartheta_{t+j} = 0$, for all $j \geq 1 \Longrightarrow \vartheta_t$ is serially uncorrelated.
- Sunspots add serial correlation to the MA of the impulse dynamics of the reduced form VARMA of a linearized DSGE model,
  1. but no cross-equations are created by the persistence tied to extrinsic shocks.
  2. This differs from adding measurement error to a linearized DSGE model, which adds cross equation restrictions to the endogenous dynamics.

Solving the Linearized NKDSGE Model: Step 4c

- The problem is the restriction $\mathbf{E}_t \vartheta_{t+j} = 0$, for all $j \geq 1$, is only necessary.
- Sims (2002) shows the necessary and sufficient conditions for the expected present discounted value formula (**SLZ**.5) to hold are

$$\sum_{j=1}^{s} \left(\mathcal{M}^{-j} \mathbb{O}_{22}^{-1} \mathbb{Q}_{2.} \mathbb{V} \zeta_{t+j} \ + \ \mathbb{Q}_{2.} \mathbb{K} \vartheta_{t+j} \right) \ = \ \mathbf{0}, \qquad (\mathbf{SLZ}.6)$$

  where s denotes a horizon sufficient for the finite forward-looking sum to hold an appropriately dimensioned column vector of zeros.
  1. The column space spanned by $\sum_{j=1}^{s} \mathcal{M}^{-j} \mathbb{O}_{22}^{-1} \mathbb{Q}_{2.} \mathbb{V}$ is a subset of the column space spanned by $Q_{2.}\mathbb{K}$.
  2. The necessary and sufficient conditions imply there is a horizon s at which information in the expected path of extrinsic shocks $\{\vartheta_{t+j}\}_{j=1}^{s}$ is 'covered' by the expected discounted path of $\{\zeta_{t+j}\}_{j=1}^{s}$ $\Longrightarrow$ the extrinsic and intrinsic shocks are linearly dependent.
- There is a solution of the multivariate first order system (**SLZ**.1) if and only if the restrictions imposed by equations (**SLZ**.6) are satisfied.
  1. The solution is $\mathbb{Q}_{2.} \mathbb{V} \zeta_{t+1} \ + \ \mathbb{Q}_{2.} \mathbb{K} \vartheta_{t+1} \ = \ \mathbf{0}$, if and only if ζ_t is serially uncorrelated, $\Longrightarrow$ only the TFP and Taylor rule innovations enter $\zeta_t \ = \ [\varepsilon_t \ \ \upsilon_t]'$.
  2. Given ζ_t is serially correlated, the restrictions of (**SLZ**.6) need to be satisfied to generate a solution for the linearized NKDSGE model (**SLZ**.1) $\Longrightarrow$ if the money growth supply rule (**NK**.8) substitutes for the Taylor rule (**NK**.9).

Solving the Linearized NKDSGE Model: Step 5a

- Given an intrinsic solution exists, Sims (2002) suggests taking some matrix $\mathbf{\Phi}$ such that $\mathbb{Q}_{1\cdot}\mathbb{K} = \mathbf{\Phi}\mathbb{Q}_{2\cdot}\mathbb{K}$.

- Premultiply equation (**SLZ**.2) with $[\mathbf{I} \quad -\mathbf{\Phi}]$, combine the result with equation (**SLZ**.3), and observe this annihilates the expectational forecast errors ϑ_t, which gives

$$\mathbb{F}_{11}\mathbb{D}_{1,t} + (\mathbb{F}_{12} - \mathbf{\Phi}\mathbb{F}_{22})\mathbb{D}_{2,t} = \mathbb{O}_{11}\mathbb{D}_{1,t-1} + (\mathbb{O}_{12} - \mathbf{\Phi}\mathbb{O}_{22})\mathbb{D}_{2,t-1} + (\mathbb{Q}_{1\cdot} - \mathbf{\Phi}\mathbb{Q}_{2\cdot})\mathbb{V}\zeta_t.$$

- Stack these equations on top of the system of equations (**SLZ**.4) to find

$$\begin{bmatrix} \mathbb{F}_{11} & \mathbb{F}_{12} - \mathbf{\Phi}\mathbb{F}_{22} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbb{D}_{1,t} \\ \mathbb{D}_{2,t} \end{bmatrix} = \begin{bmatrix} \mathbb{O}_{11} & \mathbb{O}_{12} - \mathbf{\Phi}\mathbb{O}_{22} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbb{D}_{1,t-1} \\ \mathbb{D}_{2,t-1} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbb{Q}_{1\cdot} - \mathbf{\Phi}\mathbb{Q}_{2\cdot} \\ \mathbf{0} \end{bmatrix} \mathbb{V}\zeta_t - \mathbf{E}_t\left\{ \begin{bmatrix} \mathbf{0} \\ \sum_{j=1}^{\infty} \mathcal{M}^{-j}\mathbb{O}_{22}^{-1}\mathbb{Q}_{2\cdot}\mathbb{V}\zeta_{t+j} \end{bmatrix} \right\}.$$

- The lower block has the "explosive" forward-looking variables of the linear approximate system.

Solving the Linearized NKDSGE Model: Step 5b

- This matrix system maps into the unique intrinsic solution for $\mathcal{S}_t$

$$\mathcal{S}_t = \mathbf{\Theta}_{\mathcal{S}}\mathcal{S}_{t-1} + \mathbf{\Theta}_{\zeta,0}\zeta_t + \mathbf{\Theta}_{\zeta,1}\sum_{j=1}^{\infty}\mathcal{M}^j\mathbb{O}_{22}^{-1}\mathbb{Q}_{2\cdot}\mathbb{V}\,\mathbf{E}_t\zeta_{t+j}, \qquad (\mathbf{SLZ}.7)$$

where $\mathbf{\Theta}_{\mathcal{S}} = \mathbb{Z}_{\cdot 1}\mathbb{F}_{11}^{-1}[\mathbb{O}_{11}(\mathbb{O}_{12} - \mathbf{\Phi}\mathbb{F}_{22})]\mathbb{Z}$, $\mathbf{\Theta}_{\mathbb{F}} = \begin{bmatrix}\mathbb{Z}_{\cdot 1} & \mathbb{Z}_{\cdot 2}\end{bmatrix}\begin{bmatrix}\mathbb{F}_{11}^{-1} & -\mathbb{F}_{11}^{-1}(\mathbb{F}_{12} - \mathbf{\Phi}\mathbb{F}_{22})\\ \mathbf{0} & \mathbf{I}\end{bmatrix}$,
$\mathbb{Z} = \begin{bmatrix}\mathbb{Z}_{\cdot 1} & \mathbb{Z}_{\cdot 2}\end{bmatrix}$, $\mathbf{\Theta}_{\zeta,0} = \mathbf{\Theta}_{\mathbb{F}}\cdot\begin{bmatrix}\mathbb{Q}_{1\cdot} - \mathbf{\Phi}\mathbb{Q}_{2\cdot}\\ \mathbf{0}\end{bmatrix}\cdot\mathbb{V}$, and $\mathbf{\Theta}_{\zeta,1} = -\mathbf{\Theta}_{\mathbb{F},\cdot 2}$ because $\mathbb{Z}' = \mathbb{Z}^{-1}$,
$\mathbb{Z} = \mathbb{Z}'^{-1}$, $\mathbb{Z}'_{\cdot 1}\mathbb{Z}_{\cdot 1} = \mathbf{I}$, $\mathbb{Z}'_{\cdot 2}\mathbb{Z}_{\cdot 2} = \mathbf{I}$, $\mathbb{Z}'_{\cdot 1}\mathbb{Z}_{\cdot 2} = \mathbf{0}$, and, $\mathbb{Z}'_{\cdot 1}\mathbb{Z}_{\cdot 2} = \mathbf{0}$.

- We engage the system of first-order stochastic difference equations (**SLZ**.7) to produce linear approximate solutions for the NKDSGE models.
  1. Although the system of equations (**SLZ**.1) always has a solution, is it unique? $\Longrightarrow$ Are there multiple equilibria?
  2. Uniqueness depends on the necessary and sufficient conditions involving the column space spanned by $\sum_{j=1}^{s}\mathcal{M}^{-j}\mathbb{O}_{22}^{-1}\mathbb{Q}_{2\cdot}\mathbb{V}$ being a subset of the column space spanned by $\mathbb{Q}_{2\cdot}\mathbb{K} \Longrightarrow$ the unique linear approximate solution is (**SLZ**.7); otherwise sunspot equilibria exist.
  3. If ζ_t consists only of mean zero structural innovations, $\mathbf{E}_t\zeta_{t+j} = \mathbf{0} \Longrightarrow$ the infinite horizon forward looking term in the linear approximate solution (**SLZ**.7) of the NKDSGE model disappears.

Introduction to Estimation of Linearized DSGE Models, II

- Remember a state space system yields a VAR(∞) that implies a finite order VARMA.
  1. Since a linearized DSGE model has a state space representation, apply maximum likelihood estimation (MLE) to the model's restricted VARMA.
  2. Early examples are Altuğ (1989, "Time-to-build and aggregate fluctuations: Some new evidence," *IER* 30, 889–920) and Bencivenga (1992, "An econometric study of hours and output variation with preference shocks," *IER* 33, 449–471).
  3. The MLE of the restricted VARMA produces one-step ahead forecasts that are only asymptotically normal given DSGE model shock innovations are Gaussian.
  4. The objective of the MLE is built on these forecasts $\Longrightarrow$ affects the small sample properties of the estimated DSGE model.
  5. The Kalman filter produces exact finite sample forecasts of a VARMA.

- The likelihood of several RBC models is built using the Kalman filter by Sargent (1989, "Two model of measurements and the investment accelerator," *JPE* 97, 251–287).
  1. Sargent endows each variable in the multivariate data vector of a permanent income (PI) model with idiosyncratic AR(1) measurement error processes.
  2. The several AR(1) measurement errors impose additional restrictions on the VARMA that help to identify more coefficients of the PI model.
  3. An unrestricted VAR(1) replaces the idiosyncratic AR(1) measurement errors to estimate RBC models in Ireland (2001, "Technology shocks and the business cycle: An empirical investigation," *JEDC* 25, 703–719).

- Since Sargent (1989), macroeconometrics is rife with uses of the Kalman filter, which suggests its ubiquitousness is cause for a short review.

Introduction: Forecasting ARMAs

- Since the first generation of estimated DSGE models often relied on VARMAs, let's generate intuition about linear forecasts from univariate ARMAs before constructing linear forecasts from the Kalman filter.

- The AR(1) model $x_t = \theta x_{t-1} + \epsilon_t$, $\mathbf{E}\epsilon_t = 0$ and $\mathbf{E}\epsilon_t^2 = \sigma_\epsilon^2$, yields the forecasts $\mathbf{E}_t x_{t+1} = \theta x_t$ and $\mathbf{E}_t x_{t+2} = \theta \mathbf{E}_t x_{t+1} = \theta^2 x_t$, which imply

$$\mathbf{E}_t x_{t+j} = \theta^j x_t.$$

- Observe the forecast of a AR(1) smoothly decays in an exponential fashion when $\theta \in (0, 1)$, which results in

$$\lim_{j \longrightarrow \infty} \mathbf{E}_t x_{t+j} = 0.$$

Forecasting a AR(1): Conditional and Unconditional Variances

- Since x_t is predetermined (*i.e.*, known) at date t, the conditional variance of the AR(1) of x_t also contains useful information $\text{Var}_t(x_{t+1}) = \sigma_\epsilon^2$.

- Next, apply repeated substitution to show that

$$\text{Var}_t(x_{t+2}) = \text{Var}\left(\theta^2 x_t + \epsilon_{t+1} + \theta\epsilon_t\right) = \left(1 + \theta^2\right)\sigma_\epsilon^2.$$

- One more step ahead yields the conditional variance of

$$\begin{aligned}\text{Var}_t(x_{t+3}) &= \text{Var}\left(\theta^3 x_t + \epsilon_{t+2} + \theta\epsilon_{t+1} + \theta^2\epsilon_t\right)\\ &= \left(1 + \theta^2 + \theta^4\right)\sigma_\epsilon^2.\end{aligned}$$

- By induction $\text{Var}_t(x_{t+j}) = \sum_{i=0}^{j-1}\theta^{2j}\sigma_\epsilon^2 \Longrightarrow \lim_{j\to\infty}\text{Var}_t(x_{t+j}) = \text{Var}(x_t)$, while the unconditional variance of x_t is

$$\text{Var}(x_t) = \sum_{i=0}^{\infty}\theta^{2j}\sigma_\epsilon^2 = \frac{\sigma_\epsilon^2}{1 - \theta^2}.$$

MA(q) Forecasting and Its Conditional Variance

- An important point is the first two unconditional moments of a AR(1) are the limits of the first two conditional moments.
  1. The limit of these conditional moments can be thought of as either in time $t \longrightarrow \infty$ or the forecast horizon $j \longrightarrow \infty$.
  2. This carries over to the MA(q) model.

- Consider the MA(q) model $x_t = \epsilon_t + \sum_{i=1}^{q} \phi_i \epsilon_{t-i}$.
  1. Its one-step ahead forecast is $\mathbf{E}_t x_{t+1} = \sum_{i=1}^{q} \phi_i \epsilon_{t+1-i}$ and
  2. the $j-$period ahead forecast is $\mathbf{E}_t x_{t+j} = \sum_{i=0}^{q-j} \phi_{j+i} \epsilon_{t-i}$, for $j \leq q$, where $\phi_0 \equiv 0$.

- Likewise, the conditional variance of the MA(q) is simple to compute.
  1. The conditional variance j-steps ahead is $\text{Var}_t(x_{t+j}) = \sum_{i=0}^{q-j} \phi_{j+i}^2 \sigma_\epsilon^2$ and note $\text{Var}_t(x_{t+q}) = \phi_q^2 \sigma_\epsilon^2$.
  2. The unconditional variance is $\text{Var}(x_t) = \text{Var}_t(x_t) = \sum_{i=0}^{q} \phi_i^2 \sigma_\epsilon^2$.

The State Space Form of a AR(p)

- The tricks to forecasting ARMAs consist of recognizing
  1. $\mathbf{E}_t\epsilon_{t+j} = 0$, $\text{Var}_t(\epsilon_{t+j}) = \sigma^2_\epsilon$, and to invert the AR($p$)s and ARMA($p$, q)s to produce MA(∞)s.
  2. $\Longrightarrow$ A problem is factoring the MA(∞) to compute its unknown coefficients.
- A way to avoid this problem is to write the ARMA(p, q) model in *state space* form.
- Consider the state space representation of a AR(p) model.
  1. The system of state equations is $\mathcal{S}_{t+1} = \Theta\mathcal{S}_t + \vartheta_{t+1}$, where the state vector is $\mathcal{S}_{t+1}$ and $\vartheta_t = [\epsilon_t \;\; 0 \dots 0]$ are $p \times 1$ column vectors, $\mathbf{E}\{\vartheta_t\vartheta_t'\} = \boldsymbol{\mathcal{Q}}$, and

$$\Theta = \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_{p-1} & \theta_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

  2. The system of observation equations, $\mathcal{X}_t = \mathcal{S}_t$, connects $\mathcal{S}_t$ to the observed data, where the vector of observed data is $\mathcal{X}_t = [x_t \;\; x_{t-1} \;\dots\; x_{t-p}]$, $\Longrightarrow$ implicit is an identity matrix pre-multiplies the state vector.

The State Space Representation of a ARMA(p, q)

- When $p, q > 0$, define the system of state equations

$$\mathcal{S}_{t+1} = \boldsymbol{\mathcal{A}}\mathcal{S}_t + \vartheta_{t+1}, \qquad \textbf{(SLZ.8)}$$

  where $\mathcal{S}_t$ and ϑ_t are column vectors of length $r = \max\{p, q+1\}$, and $\mathcal{A} = \Theta$, but is a $r \times r$ matrix instead of $p \times p$.

- The sample data is driven by $\mathcal{S}_{t+1}$ according to

$$\mathcal{X}_t = \Phi' \mathcal{S}_t, \qquad \textbf{(SLZ.9)}$$

  where the $r \times 1$ column vector $\Phi = [1 \;\; \phi_1 \;\; \phi_2 \;\; \ldots \;\; \phi_{r-2} \;\; \phi_{r-1}]$.

- In general, state space systems can contain exogenous or pre-determined variables.
  1. Exogenous or pre-determined variables are, for example, intercepts, deterministic time trends, and exogenous stochastic variables.
  2. White noise shocks, ω_t, can also enter the observation equations (**SLZ**.9), $\mathcal{X}_t = \Phi' \mathcal{S}_t + \omega_t$, where $\mathbf{E}\{\omega_t \omega_t'\} = \mathcal{R}$, which could be correlated with ξ_t.
  3. These shocks would be interpreted as measurement error in $\mathcal{X}_t$ and would be uncorrelated with ξ_t in this case.

The State Space Representation of a ARMA(p, q), cont.

- The ARMA(p, q) is easily recovered from the system of state and observation equations (**SLZ**.8) and (**SLZ**.9).
- The second, third, ..., and ith rows of the system of state equations (**SLZ**.8) are
  1. $\mathcal{S}_{2,t+1} = \mathcal{S}_{1,t}$, $\mathcal{S}_{3,t+1} = \mathcal{S}_{2,t} = \mathcal{S}_{1,t-1}$, $\ldots$, $\mathcal{S}_{i,t+1} = \mathcal{S}_{i-1,t} = \mathbf{L}^i \mathcal{S}_{1,t}$.
  2. $\Longrightarrow$ The first row is $\mathcal{S}_{1,t+1} = \sum_{i=1}^{r} \theta_i z_{1,t-i} + \epsilon_{t+1} \Leftrightarrow \theta(\mathbf{L}) z_{1,t+1} = \epsilon_{t+1}$, where $\theta(\mathbf{L}) = 1 - \theta_1 \mathbf{L} - \theta_2 \mathbf{L}^2 - \ldots - \theta_r \mathbf{L}^r$.
- For the ARMA(p, q), the observer equation (**SLZ**.9) gives $x_t = \phi(\mathbf{L}) z_{1,t}$, where $\phi(\mathbf{L}) = 1 + \phi_1 \mathbf{L} + \phi_2 \mathbf{L}^2 + \ldots + \phi_{r-1} \mathbf{L}^{r-1}$.
  1. Multiply both sides of the last equation by $\theta(\mathbf{L})$ to find $\theta(\mathbf{L}) x_t = \theta(\mathbf{L}) \phi(\mathbf{L}) \mathcal{S}_{1,t}$.
  2. Since $\theta(\mathbf{L}) \phi(\mathbf{L}) = \phi(\mathbf{L}) \theta(\mathbf{L})$, the result is $\theta(\mathbf{L}) x_t = \phi(\mathbf{L}) \epsilon_t$, which is the ARMA($p$, q).

The Kalman Filter: Outputs and Uses

- The system of state and observer equations define the Kalman filter (KF).
- The KF is a device for estimating the latent, hidden, or unobserved state, $\mathcal{S}_t$, of a linear time series model given conditional linear projections of the observable variables, $\mathcal{Y}_t$.
- The KF algorithm unwinds or decomposes $\mathcal{Y}_t$ into $\mathcal{S}_t$, and shock innovations to the state, ξ_t and observables, ω_t.
- The KF decomposition occurs in a sequence of linear updates of $\mathcal{S}_{t+1}$ that implies a new forecast of $\mathcal{Y}_{t+1}$, its MSE, and the MSE of $\mathcal{S}_{t+1}$ conditional on date t information.
- These forecasts are exact $\Longrightarrow$ there is no reliance on the forecast horizon going to infinity (convergence results or asymptotic theory).
- Since the KF generates forecasts of any linear times series model that can be cast in state space form, these elements are available to construct likelihood functions of ARMA models, time-varying coefficient VAR models, or almost any linear time series model.
- The classic text on the KF and filtering in general is Anderson and Moore (2005, Optimal Filtering, Mineola, NY: Dover Publications).

The Kalman Filter: Set Up

- Restate the state equation

$$\mathcal{S}_{t+1} = \boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{B}}\xi_{t+1}, \qquad (\mathbf{KF}.1)$$

where $\dim(\mathcal{S}_t) = k$, $\dim(\xi_t) = m$, $m \leq k$, $\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}' = \boldsymbol{\mathcal{Q}}$, and $\xi_{t+1} \sim \mathcal{N}(\mathbf{0}_{m\times 1}, \mathbf{I}_m)$.

- The adjustments to the observer equation are

$$\mathcal{Y}_t = \boldsymbol{\mathcal{F}}\mathcal{Z}_t + \boldsymbol{\mathcal{C}}\mathcal{S}_t + \boldsymbol{\mathcal{D}}\omega_t, \qquad (\mathbf{KF}.2)$$

where $\dim(\mathcal{Y}_t) = n$, a $p \times 1$ vector of exogenous and/or pre-determined variables, $\mathcal{Z}_t$, is added to the observations equations ($\mathcal{Z}_t$ could also be included in the state equations (**KF**.1), $\dim(\omega_t) = r$, $r \leq n$, $\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}' = \boldsymbol{\mathcal{R}}$, and $\omega_{t+1} \sim \mathcal{N}(\mathbf{0}_{r\times 1}, \mathbf{I}_r)$.

- Further, assume $\mathbf{E}\{\omega_t \xi_s'\} = 0$ for all dates t and s $\Longrightarrow$ this assumption can be relaxed with only minor changes to the Kalman filter algorithm; see Harvey (1989, section 3.2.4, pp. 112–113).

- Assume the matrices $\boldsymbol{\mathcal{A}}_{k\times k}$, $\boldsymbol{\mathcal{B}}_{k\times m}$, $\boldsymbol{\mathcal{C}}_{n\times k}$, $\boldsymbol{\mathcal{D}}_{n\times r}$, $\boldsymbol{\mathcal{Q}}_{m\times m}$, $\boldsymbol{\mathcal{R}}_{r\times r}$, and $\boldsymbol{\mathcal{F}}_{n\times p}$ are non-stochastic.

- Label the state space representation (**KF**.1) and (**KF**.2) the ABCDs state space model.

The Kalman Filter: Useful Definitions

- Define the linear projection of $\mathcal{S}_{t+1}$ on the entire histories of $\mathcal{Y}$ and $\mathcal{Z}$ from date 1 to date t and a constant as

$$\widehat{\mathcal{S}}_{t+1|t} \equiv \mathbf{E}\{\mathcal{S}_{t+1} \,|\, \mathcal{W}_t\},$$

where $\mathcal{W}_t \equiv \left[\mathcal{Y}_t' \;\; \mathcal{Y}_{t-1}' \;\; \ldots \;\; \mathcal{Y}_1' \;\; \mathcal{Z}_t' \;\; \mathcal{Z}_{t-1}' \;\; \ldots \;\; \mathcal{Z}_1'\right]'$.

1. The linearity of the Kalman filter implies $\mathbf{E}\{\cdot\}$ yields equivalent forecasts to the projection operator.
2. The Kalman filter computes $\widehat{\mathcal{S}}_{1|0}$, $\widehat{\mathcal{S}}_{2|1}$, $\ldots$, $\widehat{\mathcal{S}}_{T|T-1}$ recursively.
3. Relying on $t \longrightarrow \infty$ is unnecessary to operate the Kalman filter.

- The mean square error (MSE) of $\widehat{\mathcal{S}}_{t+1|t}$ is computed as

$$\mathbf{P}_{t+1|t} \equiv \mathbf{E}\left\{\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)'\right\}. \qquad (\mathbf{KF}.3)$$

- The analogy is to the conditional variance of $\mathcal{S}_{t+1}$.

The Kalman Filter: Initialization, I

- To start the Kalman filter, must decide how to compute $\widehat{\mathcal{S}}_{1|0}$ and $\mathbf{P}_{1|0}$.
- Since no observations on $\mathcal{Y}$ or $\mathcal{X}$ exist (other than the intercept), the best linear forecast of $\widehat{\mathcal{S}}_{1|0}$ is its unconditional expectation $\mathcal{S}_1$, $\mathbf{E}\{\mathcal{S}_1\}$.
- The corresponding MSE is

$$\mathbf{P}_{1|0} \equiv \mathbf{E}\left\{(\mathcal{S}_1 - \mathbf{E}\{\mathcal{S}_1\})(\mathcal{S}_1 - \mathbf{E}\{\mathcal{S}_1\})'\right\}.$$

- Observe from the state equations (**KF**.1) that $\mathbf{E}\{\mathcal{S}_{t+1}\} = \boldsymbol{\mathcal{A}}\mathbf{E}\{\mathcal{S}_t\}$.
- Since $\mathcal{S}_{t+1}$ is covariance stationary,

$$(\mathbf{I}_k - \boldsymbol{\mathcal{A}})\,\mathbf{E}\{\mathcal{S}_t\} = \mathbf{0}_{k\times 1}.$$

1. A unique solution yields $\mathbf{E}\{\mathcal{S}_t\} = \mathbf{0}_{k\times 1}$ because $(\mathbf{I}_k - \boldsymbol{\mathcal{A}})$ is non-singular.
2. $\Longrightarrow \widehat{\mathcal{S}}_{1|0} = \mathbf{0}_{k\times 1}$, given the specification of the system of state equations (**KF**.1).

The Kalman Filter: Initialization, II

- Note $\mathbf{P}_{1|0}$ is equivalent to the unconditional variance of $\mathcal{S} \Longrightarrow$ post-multiply (**KF**.1) by $\mathcal{S}_{t+1}$ and take the unconditional expectation through

$$\begin{aligned}\mathbf{E}\{\mathcal{S}_{t+1}\mathcal{S}'_{t+1}\} &= \mathbf{E}\left\{\left(\boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{D}}\xi_{t+1}\right)\left(\boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{D}}\xi_{t+1}\right)'\right\} \\ &= \boldsymbol{\mathcal{A}}\mathbf{E}\left\{\mathcal{S}_{t+1}\mathcal{S}'_{t+1}\right\}\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{D}}\mathbf{E}\left\{\xi_{t+1}\xi'_{t+1}\right\}\boldsymbol{\mathcal{D}}'.\end{aligned}$$

- The lack of cross-product terms results from the assumption that ξ_t and ω_t are uncorrelated at all leads and lags.
- The definition $\mathbf{E}\{\mathcal{S}_{t+1}\mathcal{S}'_{t+1}\} \equiv \boldsymbol{\Omega}_{\mathcal{S}}$ allows the last equation to be written

$$\boldsymbol{\Omega}_{\mathcal{S}} = \boldsymbol{\mathcal{A}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{Q}}.$$

- The problem is there are no easy to use numerical methods to solve this quadratic equation for the elements of $\Omega_{\mathcal{S}}$.

The Kalman Filter: Initialization, III

- This problem is overcome using the vec$(\cdot)$ operator.
- Pass the vec$(\cdot)$ operator through the last equation to find

$$\text{vec}(\mathbf{\Omega}_{S}) = \left[\mathbf{I}_{k^2} - (\mathcal{A} \bigotimes \mathcal{A})\right]^{-1} \text{vec}(\mathcal{Q}).$$

- This result follows from applying the rule $\text{vec}(\mathbf{G}_1\mathbf{G}_2\mathbf{G}_3) = (\mathbf{G}_3' \bigotimes \mathbf{G}_1)\text{vec}(\mathbf{G}_2)$ to the quadratic term, $\mathcal{A}\mathbf{\Omega}_S\mathcal{A}'$.
- The fact $\mathbf{P}_{1|0} = \Omega_S$ produces the date 1 MSE conditional on date zero information, the unconditional variance of S.
- Since the eigenvalues of $\mathcal{A}$ are inside the unit circle, the solution of $\mathbf{P}_{1|0}$ is bounded

$$\text{vec}(\mathbf{P}_{1|0}) = \left[\mathbf{I}_{k^2} - \left(\mathcal{A} \bigotimes \mathcal{A}\right)\right]^{-1} \text{vec}(\mathcal{Q}). \qquad \textbf{(KF.4)}$$

- If not, $\mathbf{P}_{1|0}$ is a free parameter of the Kalman filter; also true if $\mathbf{E}\{S_t\} \neq \mathbf{0}_{k\times 1}$.
- In this case, a guess must be made about $\mathbf{P}_{1|0}$, typically a multiple of $\mathbf{I}_{k^2}$.
$\Longrightarrow$ A diffuse prior about the unconditional variance of $\mathbf{\Omega}_S$.

The Kalman Filter: Forecast Definitions

- Given $\widehat{S}_{1|0}$ and $\mathbf{P}_{1|0}$, the goal is to compute

$$\left\{\widehat{S}_{2|1},\, \mathbf{P}_{2|1}\right\},\ \left\{\widehat{S}_{3|2},\, \mathbf{P}_{3|2}\right\},\ \ldots,\ \left\{\widehat{S}_{t|t-1},\, \mathbf{P}_{t|t-1}\right\},\ \ldots,\ \left\{\widehat{S}_{T|T-1}\ \mathbf{P}_{T|T-1}\right\}.$$

- Since the structure of the Kalman filter algorithm is the same for each date t, consider the forecasts for $\widehat{S}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$, but remember

$$\mathcal{W}_t \equiv \left[\mathcal{Y}_t'\ \ \mathcal{Y}_{t-1}'\ \ \ldots\ \ \mathcal{Y}_1'\ \ \mathcal{Z}_t'\ \ \mathcal{Z}_{t-1}'\ \ \ldots\ \ \mathcal{Z}_1'\right]'.$$

- Next, apply the expectations operator to the state equation (**KF**.1) to find

$$\mathbf{E}\left\{S_t \,\middle|\, \mathcal{W}_t\right\} = \widehat{S}_{t|t-1}.$$

- Similarly, define the conditional forecast of $\mathcal{Y}_t$ using the expectations operator

$$\widehat{\mathcal{Y}}_{t|t-1} \equiv \mathbf{E}\left\{\mathcal{Y}_t \,\middle|\, \mathcal{W}_t\right\}.$$

The Kalman Filter: Forecasting $\mathcal{Y}_t$

- The system of observer equations (**KF**.2) also yield

$$\mathbf{E}\Big\{\mathcal{Y}_t \,|\, \mathcal{Z}_t,\, \mathcal{S}_t\Big\} \;=\; \boldsymbol{\mathcal{F}}\mathcal{Z}_t \;+\; \boldsymbol{C}\mathcal{S}_t.$$

- The KF needs to forecast $\widehat{\mathcal{Y}}_{t|t-1}$, which is constructed using
  1. a forecast of $\mathcal{S}_t$ computed on the state equations (**KF**.1) given $\mathcal{S}_t \Longrightarrow \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{S}_t\Big\}$,
  2. and by applying the law of iterated expectations to this forecast after conditioning on $\mathcal{Y}_t$ and $\mathcal{Z}_t$

$$\mathbf{E}\left\{\mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{S}_t\Big\} \,|\, \mathcal{Y}_t,\, \mathcal{Z}_t\right\} \;=\; \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{Y}_t,\, \mathcal{Z}_t\Big\} \;=\; \widehat{\mathcal{S}}_{t|t-1}.$$

- Steps 1 and 2 imply the exact forecast of $\mathcal{Y}_t$ is

$$\widehat{\mathcal{Y}}_{t|t-1} \;=\; \boldsymbol{\mathcal{F}}\mathcal{Z}_t \;+\; \boldsymbol{C}\widehat{\mathcal{S}}_{t|t-1}, \qquad (\mathbf{KF}.5)$$

  1. which relies on the known coefficient matrices $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{C}$, the predetermined and deterministic variables in the vector $\mathcal{Z}_t$, and the forecast $\widehat{\mathcal{S}}_{t|t-1}$.
  2. $\Longrightarrow$ The forecast $\widehat{\mathcal{Y}}_{t|t-1}$ is exact because it is grounded on population moments, known coefficient matrices, and predetermined and deterministic variables.

The Kalman Filter: Forecasting $\mathcal{Y}_t$, cont.

- The observer equations (**KF**.2) and the observer forecast generating equations (**KF**.5) are responsible for the system of observer forecast errors

$$\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1} = \boldsymbol{\mathcal{F}} \mathcal{Z}_t + \boldsymbol{C} \mathcal{S}_t + \omega_t - \left(\boldsymbol{\mathcal{F}} \mathcal{Z}_t + \boldsymbol{C} \widehat{\mathcal{S}}_{t|t-1} \right) = \boldsymbol{C} \left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t-1} \right) + \omega_t . \quad (\textbf{KF}.6)$$

- The system of forecast errors (**KF**.6) produces the MSE of forecasts of $\mathcal{Y}_t$

$$\mathbf{E}\left\{ \left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1} \right) \left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1} \right)' \right\} = \mathbf{E}\left\{ \boldsymbol{C} \left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t-1} \right) \left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t-1} \right)' \boldsymbol{C}' + \omega_t \omega_t' \right\}, \quad (\textbf{KF}.7)$$

  where cross-product terms disappear because $\mathbf{E}\left\{ \omega_t \left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t-1} \right)' \right\} = \mathbf{0}$.

- Construct the MSE of $\mathcal{Y}_t$'s forecast errors by substituting the MSE of $\widehat{\mathcal{S}}_{t|t-1}$, equations (**KF**.3), and $\mathbf{E}\left\{ \omega_t \omega_{t+j}' \right\} = \boldsymbol{\mathcal{R}}$ into the right hand side of equations (**KF**.7) to find

$$\mathbf{E}\left\{ \left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1} \right) \left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1} \right)' \right\} = \boldsymbol{C} \mathbf{P}_{t|t-1} \boldsymbol{C}' + \boldsymbol{\mathcal{R}}, \quad (\textbf{KF}.8)$$

  given only the known coefficient matrices $\boldsymbol{C}$ and $\boldsymbol{\mathcal{R}}$ and the unknown MSE, $\mathbf{P}_{t+1|t}$.

The Kalman Filter: Forecasting $\mathcal{S}_t$

- The system of equations (**KF**.5) need $\widehat{\mathcal{S}}_{t|t-1}$, to generate the forecast $\widehat{\mathcal{Y}}_{t|t-1}$.
- Besides an exact forecast of $\mathcal{S}_t$ conditional on date $t-1$ information, $\mathcal{W}_{t-1}$, the KF yields an exact forecast of $\mathcal{S}_t$ conditional on $\mathcal{W}_t$

$$\widehat{\mathcal{S}}_{t|t} \;=\; \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{Y}_t,\; \mathcal{Y}_{t-1}, \ldots, \mathcal{Z}_t,\; \mathcal{Z}_{t-1}, \ldots\Big\} \;=\; \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{Y}_t\Big\}.$$

- The previous results depends on the rule for updating a linear projection, which is discussed in Hamilton (1994, sections 4.1, pp. 72–77 and 4.4, pp. 92–100).
  1. At date t, the best forecasts of $\mathcal{Y}$ and $\mathcal{Z}$ are the realizations of these objects at date t, $\mathcal{Y}_t$ and $\mathcal{Z}_t$.
  2. At date t, the best linear forecast of the latent state vector, $\mathcal{S}_t$, consists of the forecast at date $t-1$, $\widehat{\mathcal{S}}_{t|t-1}$, and the uncertainty that surrounds $\widehat{\mathcal{S}}_{t|t}$.
  3. The uncertainty surrounding the forecast $\widehat{\mathcal{S}}_{t|t}$ is comprised of the forecast errors $\mathcal{S}_t \;-\; \widehat{\mathcal{S}}_{t|t-1}$ and $\mathcal{Y}_t \;-\; \widehat{\mathcal{Y}}_{t|t-1}$.

The Kalman Filter: Forecasting S_t, cont.

- Using the the rule for updating a linear projection gives

$$\widehat{S}_{t|t} = \widehat{S}_{t|t-1} + \mathbf{E}\left\{\left(S_t - \widehat{S}_{t|t-1}\right)\left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1}\right)'\right\}$$
$$\times \left[\mathbf{E}\left\{\left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1}\right)\left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1}\right)'\right\}\right]^{-1}\left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1}\right). \qquad (\mathbf{KF}.9)$$

- The MSE terms represent the "LS" estimate of the response of the date t forecast error of S_t to the date t forecast error of $\mathcal{Y}_t$ conditional on date $t-1$ information.

- The first MSE term of equations (**KF**.9) is computed with the MSE of $\mathcal{Y}_t$ conditional on date $t-1$ information, equations (**KF**.8)

$$\begin{aligned}\mathbf{E}\left\{\left(S_t - \widehat{S}_{t|t-1}\right)\left(\mathcal{Y}_t - \widehat{\mathcal{Y}}_{t|t-1}\right)'\right\} &= \mathbf{E}\left\{\left(S_t - \widehat{S}_{t|t-1}\right)\left[\boldsymbol{C}\left(S_t - \widehat{S}_{t|t-1}\right) + \omega_t\right]'\right\} \\ &= \mathbf{E}\left\{\left(S_t - \widehat{S}_{t|t-1}\right)\left(S_t - \widehat{S}_{t|t-1}\right)'\right\}\boldsymbol{C}' \\ &= \mathbf{P}_{t|t-1}\boldsymbol{C}',\end{aligned}$$

where the last equality follows from the MSE of $\widehat{S}_{t|t-1}$ found in equations (**KF**.3).

The Kalman Filter: Forecasting $\mathcal{S}_t$, cont.

- Finally, substitute for the forecast of $\mathcal{Y}_t$ conditional on date $t-1$ information, the equations (**KF**.5), and its associated MSE, equations (**KF**.8),

$$\widehat{\mathcal{S}}_{t|t} = \widehat{\mathcal{S}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathcal{C}'\left[\mathcal{C}\mathbf{P}_{t|t-1}\mathcal{C}' + \mathcal{R}\right]^{-1}\left(\mathcal{Y}_t - \mathcal{F}\mathcal{Z}_t - \mathcal{C}\widehat{\mathcal{S}}_{t|t-1}\right). \qquad (\mathbf{KF}.10)$$

  which is exact because it is grounded on population moments, known coefficient matrices, and predetermined and deterministic variables.

- The definition of the MSE of $\widehat{\mathcal{S}}_{t|t}$ is $\mathbf{P}_{t|t} \equiv \mathbf{E}\left\{\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)'\right\}$, which can be computed employing similar tricks to produce

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathcal{C}'\left[\mathcal{C}\mathbf{P}_{t|t-1}\mathcal{C}' + \mathcal{R}\right]^{-1}\mathcal{C}\mathbf{P}_{t|t-1}. \qquad (\mathbf{KF}.11)$$

- Notice the MSE of $\widehat{\mathcal{S}}_{t|t}$, $\mathbf{P}_{t|t}$, equals the update of $\mathbf{P}_{t|t-1}$ plus the MSE of the forecast error of $\mathcal{Y}_t$; see the updating equations (**KF**.10) of $\widehat{\mathcal{S}}_{t|t}$.

The Kalman Filter: Forecasting $\mathcal{S}_t$, cont.

- The definition $\widehat{\mathcal{S}}_{t+1|t} \equiv \mathbf{E}\left\{\mathcal{S}_{t+1} \,|\, \mathcal{Y}_t\right\}$ and the system of state equations (**KF**.1) yield

$$\widehat{\mathcal{S}}_{t+1|t} = \boldsymbol{\mathcal{A}}\mathbf{E}\left\{\mathcal{S}_{t+1} \,|\, \mathcal{Y}_t\right\} + \boldsymbol{\mathcal{B}}\mathbf{E}\left\{\xi_{t+1} \,|\, \mathcal{Y}_t\right\} = \boldsymbol{\mathcal{A}}\widehat{\mathcal{S}}_{t|t}.$$

- Replace $\widehat{\mathcal{S}}_{t|t}$ with its update given in equation (**KF**.10) to find

$$\widehat{\mathcal{S}}_{t+1|t} = \boldsymbol{\mathcal{A}}\widehat{\mathcal{S}}_{t|t-1} + \mathcal{K}_t\left(\mathcal{Y}_t - \boldsymbol{\mathcal{F}}\mathcal{Z}_t - \boldsymbol{\mathcal{C}}\widehat{\mathcal{S}}_{t|t-1}\right), \qquad (\mathbf{KF}.12)$$

where $\mathcal{K}_t \equiv \boldsymbol{\mathcal{A}}\mathbf{P}_{t|t-1}\boldsymbol{\mathcal{C}}'\left[\boldsymbol{\mathcal{C}}\mathbf{P}_{t|t-1}\boldsymbol{\mathcal{C}}' + \boldsymbol{\mathcal{R}}\right]^{-1}$.

- Equation (**KF**.12) provides the forecast of $\mathcal{S}_{t+1}$ conditional on date t information given the known coefficient matrices $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{C}}$, and $\boldsymbol{\mathcal{R}}$, the unknown MSE $\mathbf{P}_{t|t-1}$, and the "innovation" in $\mathcal{Y}_t$.

- Once $\mathbf{P}_{t|t-1}$ is computed, $\widehat{\mathcal{Y}}_{t+1|t}$ is constructed with equation (**KF**.5).

The Kalman Filter: Forecasting $\mathcal{S}_t$, finis

- Calculating the MSE of $\widehat{\mathcal{S}}_{t+1|t}$, equation (**KF**.3), starts with the fact $\widehat{\mathcal{S}}_{t+1|t} = \boldsymbol{\mathcal{A}}\widehat{\mathcal{S}}_{t+1|t} \Longrightarrow$

$$\begin{aligned}\mathbf{P}_{t+1|t} &= \mathbf{E}\left\{\left(\boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{B}}\xi_t - \boldsymbol{\mathcal{A}}\mathcal{S}_{t|t}\right)\left(\boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{B}}\xi_t - \boldsymbol{\mathcal{A}}\mathcal{S}_{t|t}\right)'\right\} \\ &= \boldsymbol{\mathcal{A}}\mathbf{E}\left\{\left(\mathcal{S}_t - \mathcal{S}_{t|t}\right)\left(\mathcal{S}_t - \mathcal{S}_{t|t}\right)'\right\}\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{B}}\mathbf{E}\left\{\xi_t\xi_t'\right\}\boldsymbol{\mathcal{B}}', \\ &= \boldsymbol{\mathcal{A}}\mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{Q}}. \qquad (\mathbf{KF}.13)\end{aligned}$$

- Recover the updating equation for the MSE of $\mathcal{S}_{t+1}$ conditional on date t information by substituting $\mathbf{P}_{t|t}$ from (**KF**.11) into (**KF**.13)

$$\mathbf{P}_{t+1|t} = \boldsymbol{\mathcal{A}}\mathbf{P}_{t|t-1}\boldsymbol{\mathcal{A}}' - \mathcal{K}_t\boldsymbol{\mathcal{C}}\mathbf{P}_{t|t-1}\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{Q}}. \qquad (\mathbf{KF}.14)$$

- Given $\mathcal{S}_1$ and $\mathbf{P}_{1|0}$, all the objects on the right hand sides of (**KF**.12) and (**KF**.14) are known with either certainty or are already computed recursively by the Kalman filter algorithm.

The Kalman Filter: Odds and Ends

- The KF is a device for estimating the latent state vector of a macro model $\Longrightarrow$ the hidden state vector of an aggregate economy.
- These notes are just one of many ways to define, express, and solve the KF to accomplish this task.
- This approach is among the most useful ways to learn the KF algorithm, especially for many time series model employed by macroeconomists.
- When the KF is connected to a macro model, the unobserved or latent state vector S_{t+1} may possess a structural interpretation.
  1. This suggests there is interest in estimates of $\widehat{S}_{t+1|T} \Longrightarrow$ the estimate of S_{t+1} conditional on the entire sample from date $t = 1$ through date T.
  2. Reverse or backward filtering recursions yield a sequence of estimates of $\widehat{S}_{t+1|T}$ and associated MSEs $\Longrightarrow$ Kalman smoothing; see Hamilton (1994, section 13.5, pp. 394–397).

A Brief Review of the Kalman Filter

- The KF produces estimates of the state vector, $\mathcal{S}_{t+1}$, $\widehat{\mathcal{S}}_{t+1|t} \equiv \mathbf{E}\left\{\mathcal{S}_{t+1} \,|\, \mathcal{W}_t\right\}$, and its MSE, $\mathbf{P}_{t+1|t} \equiv \mathbf{E}\left\{\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)'\right\}$.
- The recursions generating these estimates are

$$\widehat{\mathcal{S}}_{t|t} = \widehat{\mathcal{S}}_{t|t-1} + \mathbf{P}_{t|t-1}\boldsymbol{\mathcal{C}}'\left[\boldsymbol{\mathcal{C}}\mathbf{P}_{t|t-1}\boldsymbol{\mathcal{C}}' + \boldsymbol{\mathcal{R}}\right]^{-1}\left(\mathcal{Y}_t - \boldsymbol{\mathcal{F}}\mathcal{Z}_t - \boldsymbol{\mathcal{C}}\widehat{\mathcal{S}}_{t|t-1}\right), \qquad \text{(KF.15)}$$

$$\widehat{\mathcal{S}}_{t+1|t} = \boldsymbol{\mathcal{A}}\widehat{\mathcal{S}}_{t|t}, \qquad \text{(KF.16)}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\boldsymbol{\mathcal{C}}'\left[\boldsymbol{\mathcal{C}}\mathbf{P}_{t|t-1}\boldsymbol{\mathcal{C}}' + \boldsymbol{\mathcal{R}}\right]^{-1}\boldsymbol{\mathcal{C}}\mathbf{P}_{t|t-1}, \qquad \text{(KF.17)}$$

$$\mathbf{P}_{t+1|t} = \boldsymbol{\mathcal{A}}\mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{Q}}. \qquad \text{(KF.18)}$$

- Estimates of $\widehat{\mathcal{S}}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$ are conditional on $\mathcal{W}_t \equiv \left[\mathcal{Y}_t'\ \mathcal{Y}_{t-1}' \ldots \mathcal{Y}_1'\ \mathcal{Z}_t'\ \mathcal{Z}_{t-1}' \ldots \mathcal{Z}_1'\right]'$.
    1. Estimates grounded only in data from date 1 to date t are useful as "real time" (in-sample) forecasts or as generated regressors.
    2. Examples are "real time" estimates of the hidden state of the business cycle or, say, regress sample output growth on lags of estimated TFP growth to test predictive information of the latter for the former.
- The KF estimates are not the model's deep structural estimates.
    1. Structural estimates rely on the entire sample $\Longrightarrow$ the Kalman smoother (KS).
    2. The KS produces two-sided estimates of $\mathcal{S}_t$ $\Longrightarrow$ run the KF forward from $t = 1$ to T and the smoother backwards from date T to $t = 1$ $\Longrightarrow$ $\widehat{\mathcal{S}}_{t|T}$ and $\mathbf{P}_{t|T}$.

The Kalman Smoother Algorithm, I

- A goal of the KS is to construct $\left\{\widehat{\mathcal{S}}_{t|T}\right\}_{t=1}^{T}$ by estimating $\widehat{\mathcal{S}}_{t|t+1}$, given knowledge of $\mathcal{S}_{t+1} \Longrightarrow$ the linear projection of $\mathcal{S}_t$ on $\mathcal{S}_{t+1}$ and $\mathcal{W}_t$

$$\begin{aligned}\mathbf{E}\left\{\mathcal{S}_t \,|\, \mathcal{S}_{t+1},\, \mathcal{W}_t\right\} \;=\; & \widehat{\mathcal{S}}_{t|t} + \mathbf{E}\left\{\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)'\right\} \\ & \times \left[\mathbf{E}\left\{\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)'\right\}\right]^{-1}\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right). \qquad (\mathbf{KS}.1)\end{aligned}$$

- Since $\mathbf{E}\left\{\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right)'\right\} = \mathbf{E}\left\{\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)\left(\boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{B}}\xi_{t+1} - \boldsymbol{\mathcal{A}}\widehat{\mathcal{S}}_{t|t}\right)'\right\}$ and $\mathbf{E}\left\{\mathcal{S}_t\xi_{t+1}\right\} = \mathbf{E}\left\{\mathcal{S}_{t|t}\xi_{t+1}\right\} = \mathbf{0}$, $\mathbf{E}\left\{\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)\left(\boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{B}}\xi_{t+1} - \boldsymbol{\mathcal{A}}\widehat{\mathcal{S}}_{t|t}\right)'\right\} =$ $\mathbf{E}\left\{\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)\left(\mathcal{S}_t - \widehat{\mathcal{S}}_{t|t}\right)'\right\}\boldsymbol{\mathcal{A}}' = \mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}'$, which after substitution into (**KS**.1) yields the exact forecast

$$\begin{aligned}\mathbf{E}\left\{\mathcal{S}_t \,|\, \mathcal{S}_{t+1},\, \mathcal{W}_t\right\} &= \widehat{\mathcal{S}}_{t|t} + \mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}'\mathbf{P}^{-1}_{t+1|t}\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right), \\ &= \widehat{\mathcal{S}}_{t|t} + \boldsymbol{\mathcal{H}}_t\left(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\right), \qquad (\mathbf{KS}.2)\end{aligned}$$

where $\boldsymbol{\mathcal{H}}_t = \mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}'\mathbf{P}^{-1}_{t+1|t}$.

The Kalman Smoother Algorithm, II

- By induction, $\mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{S}_{t+1},\ \mathcal{W}_t\Big\} = \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{S}_{t+1},\ \mathcal{W}_T\Big\} \Longrightarrow$ the additional information in $\mathcal{Y}_{t+j}$ and $\mathcal{Z}_{t+j}$, $j \geq 1$, is already contained in $\mathcal{S}_{t+1}$ or $\mathcal{S}_t - \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{S}_{t+1},\ \mathcal{W}_t\Big\} \perp \mathcal{S}_{t+1}$.
- However, conditioning the state only on $\mathcal{W}_T$, $\mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{W}_T\Big\}$, results in

$$\begin{aligned} \mathbf{E}\Big\{\mathcal{S}_t \,|\, \mathcal{W}_T\Big\} &= \widehat{\mathcal{S}}_{t|t} + \mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}'\mathbf{P}^{-1}_{t+1|t}\left(\mathbf{E}\Big\{\mathcal{S}_{t+1} \,|\, \mathcal{W}_T\Big\} - \widehat{\mathcal{S}}_{t+1|t}\right), \\ \widehat{\mathcal{S}}_{t|T} &= \widehat{\mathcal{S}}_{t|t} + \boldsymbol{\mathcal{H}}_t\left(\widehat{\mathcal{S}}_{t+1|T} - \widehat{\mathcal{S}}_{t+1|t}\right), \end{aligned} \qquad (\mathbf{KS}.3)$$

  where the exact forecasts $\widehat{\mathcal{S}}_{t|t}$ and $\widehat{\mathcal{S}}_{t+1|t}$ are untouched depending only on information up to and including date t.
- First run the KF recursions using equations (**KF**.15), (**KF**.16), (**KF**.17), and (**KF**.18) to construct $\Big\{\widehat{\mathcal{S}}_{t|t},\ \ \widehat{\mathcal{S}}_{t+1|t},\ \ \mathbf{P}_{t|t},\ \ \mathbf{P}_{t+1|t}\Big\}_{t=1}^{T}$.
  1. The last element of $\Big\{\widehat{\mathcal{S}}_{t|t}\Big\}_{t=1}^{T}$ is equated to the smoothed quantity $\widehat{\mathcal{S}}_{T|T}$, and
  2. engage the last two sequences to create $\Big\{\boldsymbol{\mathcal{H}}_t\Big\}_{t=1}^{T-1}$ using $\boldsymbol{\mathcal{H}}_t = \mathbf{P}_{t|t}\boldsymbol{\mathcal{A}}'\mathbf{P}^{-1}_{t+1|t}$.
  3. $\Longrightarrow$ At date $T-1$, compute $\widehat{\mathcal{S}}_{T-1|T} = \widehat{\mathcal{S}}_{T-1|T-1} + \boldsymbol{\mathcal{H}}_{T-1}\left(\widehat{\mathcal{S}}_{T|T} - \widehat{\mathcal{S}}_{T|T-1}\right)$, where the KF is the source of $\widehat{\mathcal{S}}_{T-1|T-1}$, $\widehat{\mathcal{S}}_{T|T-1}$, and $\boldsymbol{\mathcal{H}}_t$.
- Repeat this process from dates $t = T-2, \ldots, 1$ to finish constructing $\Big\{\widehat{\mathcal{S}}_{t|T}\Big\}_{t=1}^{T}$, where the relevant recursion is the system of equations (**KS**.3).

The Kalman Smoother Algorithm, III

- The KS also produces sequences of the MSEs of $\left\{\widehat{S}_{t|T}\right\}_{t=1}^{T}$.
- The MSE, $\mathbf{P}_{t|T}$, is constructed in several steps.
  1. Add S_t to each side the system of equations (**KS**.3) after multiplying by -1

$$\begin{aligned} S_t - \widehat{S}_{t|T} &= S_t - \widehat{S}_{t|t} - \boldsymbol{\mathcal{H}}_t\left(\widehat{S}_{t+1|T} - \widehat{S}_{t+1|t}\right), \\ S_t - \widehat{S}_{t|T} + \boldsymbol{\mathcal{H}}_t\widehat{S}_{t+1|T} &= S_t - \widehat{S}_{t|t} + \boldsymbol{\mathcal{H}}_t\widehat{S}_{t+1|t}. \end{aligned}$$

  2. Multiple the latter system of equations by its transpose

$$\begin{aligned} &\mathbf{E}\left\{\left(S_t - \widehat{S}_{t|T}\right)\left(S_t - \widehat{S}_{t|T}\right)'\right\} + \boldsymbol{\mathcal{H}}_t\mathbf{E}\left\{\widehat{S}_{t+1|T}\widehat{S}'_{t+1|T}\right\}\boldsymbol{\mathcal{H}}'_t \\ &\quad = \mathbf{E}\left\{\left(S_t - \widehat{S}_{t|t}\right)\left(S_t - \widehat{S}_{t|t}\right)'\right\} + \boldsymbol{\mathcal{H}}_t\mathbf{E}\left\{\widehat{S}_{t+1|t}\widehat{S}'_{t+1|t}\right\}\boldsymbol{\mathcal{H}}'_t, \end{aligned}$$

  3. where $\widehat{S}_{t+1|T} \perp S_t - \widehat{S}_{t|T}$ and $\widehat{S}_{t+1|t} \perp S_t - \widehat{S}_{t|t}$ explain the absence of cross-product terms.

The Kalman Smoother Algorithm, IV

- The next step recognizes the last equality gives

$$\mathbf{P}_{t|T} \;=\; \mathbf{P}_{t|t} \;+\; \boldsymbol{\mathcal{H}}_t \mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1|t}\widehat{\mathcal{S}}'_{t+1|t} \;-\; \widehat{\mathcal{S}}_{t+1|T}\widehat{\mathcal{S}}'_{t+1|T}\Big\}\boldsymbol{\mathcal{H}}'_t. \tag{KS.4}$$

- The difference of covariance matrices $\mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1|t}\widehat{\mathcal{S}}'_{t+1|t} - \widehat{\mathcal{S}}_{t+1|T}\widehat{\mathcal{S}}'_{t+1|T}\Big\}$ is evaluated by adding and subtracting $\mathbf{E}\Big\{\mathcal{S}_{t+1}\mathcal{S}'_{t+1}\Big\}$

$$\Big(\mathbf{E}\Big\{\mathcal{S}_{t+1}\mathcal{S}'_{t+1}\Big\} - \mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1|T}\widehat{\mathcal{S}}'_{t+1|T}\Big\}\Big) - \Big(\mathbf{E}\Big\{\mathcal{S}_{t+1}\mathcal{S}'_{t+1}\Big\} - \mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1|t}\widehat{\mathcal{S}}'_{t+1|t}\Big\}\Big)$$

$$= \mathbf{E}\left\{\Big(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|T}\Big)\Big(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|T}\Big)'\right\} - \mathbf{E}\left\{\Big(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\Big)\Big(\mathcal{S}_{t+1} - \widehat{\mathcal{S}}_{t+1|t}\Big)'\right\}.$$

1. where the cross-products $\mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1}\widehat{\mathcal{S}}'_{t+1|T}\Big\} = \mathbf{E}\left\{\Big(\widehat{\mathcal{S}}_{t+1} - \widehat{\mathcal{S}}'_{t+1|T} + \widehat{\mathcal{S}}'_{t+1|T}\Big)\widehat{\mathcal{S}}'_{t+1|T}\right\}$ $= \mathbf{E}\left\{\Big(\widehat{\mathcal{S}}_{t+1} - \widehat{\mathcal{S}}'_{t+1|T}\Big)\widehat{\mathcal{S}}'_{t+1|T}\right\} + \mathbf{E}\left\{\widehat{\mathcal{S}}'_{t+1|T}\widehat{\mathcal{S}}'_{t+1|T}\right\} = \mathbf{E}\left\{\widehat{\mathcal{S}}'_{t+1|T}\widehat{\mathcal{S}}'_{t+1|T}\right\}$ and
2. $\mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1}\widehat{\mathcal{S}}'_{t+1|t}\Big\} = \mathbf{E}\left\{\Big(\widehat{\mathcal{S}}_{t+1} - \widehat{\mathcal{S}}'_{t+1|t} + \widehat{\mathcal{S}}'_{t+1|t}\Big)\widehat{\mathcal{S}}'_{t+1|t}\right\} = \mathbf{E}\left\{\Big(\widehat{\mathcal{S}}_{t+1} - \widehat{\mathcal{S}}'_{t+1|t}\Big)\widehat{\mathcal{S}}'_{t+1|t}\right\}$ $+ \mathbf{E}\left\{\widehat{\mathcal{S}}'_{t+1|t}\widehat{\mathcal{S}}'_{t+1|t}\right\} = \mathbf{E}\left\{\widehat{\mathcal{S}}'_{t+1|t}\widehat{\mathcal{S}}'_{t+1|t}\right\}$.
3. $\Longrightarrow \mathbf{E}\Big\{\widehat{\mathcal{S}}_{t+1|t}\widehat{\mathcal{S}}'_{t+1|t} - \widehat{\mathcal{S}}_{t+1|T}\widehat{\mathcal{S}}'_{t+1|T}\Big\} = \mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}$.

The Kalman Smoother Algorithm, V

- The KS algorithm finishes by substituting

$$\mathbf{E}\left\{\widehat{S}_{t+1|t}\widehat{S}'_{t+1|t} - \widehat{S}_{t+1|T}\widehat{S}'_{t+1|T}\right\} = \mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}$$

into the system of equations (**KS**.4) to obtain

$$\mathbf{P}_{t|T} \ = \ \mathbf{P}_{t|t} \ + \ \mathcal{H}_t\left(\mathbf{P}_{t+1|T} \ - \ \mathbf{P}_{t+1|t}\right)\mathcal{H}'_t. \qquad (\mathbf{KS}.5)$$

- The system of equations (**KS**.5) is the updating recursion of $\mathbf{P}_{t|T}$, which generates the sequence $\left\{\mathbf{P}_{t|T}\right\}_{t=1}^{T}$.
- This completes the KS algorithm in which the two-sided smoothed estimates of $\left\{\widehat{S}_{t|T}\right\}_{t=1}^{T}$ and $\left\{\mathbf{P}_{t|T}\right\}_{t=1}^{T}$ are computed using the recursions (**KS**.3) and (**KS**.5), respectively.

The Kalman Filter: Parameter Estimation

- The KF also provides a mechanism for estimating the elements of the coefficient matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{F}$, $\mathcal{Q}$, and $\mathcal{R}$.

- The assumptions $\mathcal{S}_1$ and $\{\omega_t,\ \xi_t\}_{t=1}^{T}$ are drawn from a multivariate Gaussian distribution imply

  1. the forecasts $\left\{\widehat{\mathcal{S}}_{t|t-1},\ \mathcal{Y}_{t|t-1}\right\}_{t=1}^{T}$ are optimal conditional on $\mathcal{Z}_t$ and $\mathcal{Y}_{t-1}$, and
  2. $\mathcal{Y}_t$ is Gaussian conditional on $\mathcal{Z}_t$ and $\mathcal{Y}_{t-1}$ with mean and variance computed in (**KF**.5) and (**KF**.8) $\Longrightarrow \mathcal{Y}_t \,\Big|\, \mathcal{Z}_t,\, \mathcal{Y}_{t-1} \sim \mathcal{N}\left(\mathcal{F}\mathcal{Z}_t + \mathcal{C}\widehat{\mathcal{S}}_{t|t-1},\ \mathcal{C}\mathbf{P}_{t|t-1}\mathcal{C}' + \mathcal{R}\right)$.
  3. $\Longrightarrow$ The (natural) log of the conditional joint density function of $\mathcal{Y}_t$ is

$$\ln\left[f_{\mathcal{Y}_t\,|\,\mathcal{Z}_t,\mathcal{Y}_{t-1}}\left(\mathcal{Y}_t \,\Big|\, \mathcal{Z}_t,\, \mathcal{Y}_{t-1}\right)\right] = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\left|\mathcal{C}\mathbf{P}_{t|t-1}\mathcal{C}' + \mathcal{R}\right|$$
$$- \frac{1}{2}\left(\mathcal{Y}_t - \mathcal{F}\mathcal{Z}_t - \mathcal{C}\widehat{\mathcal{S}}_{t|t-1}\right)'\left[\mathcal{C}\mathbf{P}_{t|t-1}\mathcal{C}' + \mathcal{R}\right]^{-1}\left(\mathcal{Y}_t - \mathcal{F}\mathcal{Z}_t - \mathcal{C}\widehat{\mathcal{S}}_{t|t-1}\right). \quad (\textbf{KS}.6)$$

- Compute the sample log likelihood by summing (**KS**.6) across all dates t

$$\sum_{t=1}^{T}\ln\left[f_{\mathcal{Y}_t\,|\,\mathcal{Z}_t,\mathcal{Y}_{t-1}}\left(\mathcal{Y}_t \,\Big|\, \mathcal{Z}_t,\, \mathcal{Y}_{t-1}\right)\right]. \quad (\textbf{KS}.7)$$

The Kalman Filter: Parameter Estimation, cont.

- There are several implementation issue involved with estimating

1. $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{F}$, $\mathcal{Q}$, and $\mathcal{R}$ by maximizing the sample log likelihood (**KS**.7) given log of the conditional joint density function of $\mathcal{Y}_t$, which is equation (**KS**.6).
2. Identification of the coefficients of interest is difficult; see the example of a MA(p) in Hamilton (1994, pp. 387–388).
3. When the $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{F}$, $\mathcal{Q}$, and $\mathcal{R}$ matrices are estimated by ML, uncertainty about the smoothed estimates of $\mathcal{S}_{t+1}$, $\widehat{\mathcal{S}}_{t+1|T}$, exists.
4. Although the MSE of the smoothed estimate $\widehat{\mathcal{S}}_{t+1|T}$ (*i.e.*, the uncertainty associated with the KF algorithm) is available, there exists uncertainty in the smoother because of the sampling uncertainty of the coefficient estimates.
5. The sampling uncertainty of the MLE requires Monte Carlo or simulation methods necessary to quantify the uncertainty of the MSE of estimates of $\mathcal{S}_t$ produced by the KF and/or KS.
6. It is often not reasonable to assume that the shock innovations ξ_t and ω_t are drawn from a Gaussian distribution $\Longrightarrow$ MLE remain asymptotically consistent, but not asymptotically efficient.
7. Use quasi-MLE methods to construct an estimator of the covariance matrix of the coefficient estimates; see Hamilton (1994, section 13.5 p. 389).

Introduction to Bayesian Estimation of DSGE Models

- There are several reasons Bayesian times series methods are often preferred by macroeconomists studying DSGE models.
    1. Advances in Bayesian theory give macroeconomists a greater array of tools to engage when estimating and evaluating DSGE models.
    2. Computational power has shrunk the time necessary to estimate and evaluate DSGE models using Markov chain Monte Carlo (MCMC) simulators.
    3. DSGE models often pose identification problems for frequentist estimators that no amount of data and computing power can overcome.
- Macroeconomists find Bayesian estimation and evaluation methods useful because DSGE models are often seen as abstractions of actual economies.
- A frequentist econometrician might say that DSGE models are misspecified versions of the true model.
    1. This approach conflicts with views macroeconomists often hold about DSGE models $\Longrightarrow$ the mantra is "All models are false."
    2. Since Bayesians deny the existence of a true model, employing Bayesian tools to study DSGE models matches the views held by many macroeconomists.

Introduction: Frequentist Estimation of DSGE Models

- DSGE models have been estimated using frequentist methods since the late 1980s.

- Frequentist estimation of DSGE models rely on maximum likelihood estimation (MLE), generalized methods of moments (GMM), or indirect inference (II).

  1. MLE of the restricted VARMA of the linearized DSGE model is carried out by Altuğ (1989) and Bencivenga (1992).
  2. Christiano and Eichenbaum (1992, "Current real-business-cycle theories and aggregate labor-market fluctuations," *Ameican Economic Review* 82, 430–450) estimate a RBC model's parameters using GMM.
  3. II estimation of linearized DSGE models Macro is developed by Smith (1993, "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics* 18, S63–S84 and Gourieroux, Monfort, and Renault (1993, "Indirect Inference," *Journal of Applied Econometrics* 18, S85–S118)

- A recent survey of the GMM estimator applied to DSGE model is Ruge-Murcia (2013, "Generalized method of moments estimation of DSGE models," in Handbook of Research Methods and Applications on Empirical Macroeconomics, N. Hashimzade and M. Thornton (eds.), Edward Elgar: Cheltenham, UK).

- Dridi, Guay, and Renault (2007, "Indirect inference and calibration of dynamic stochastic general equilibrium models," *Journal of Econometrics*, 136, 397–430) is a useful update of the II estimator employed to linearized DSGE models.

Introduction: Identification of DSGE Models, I

- Identification matters for ML estimation of DSGE models.
    1. Altuğ (1989), Bencivenga (1992), and Ireland (2001) adopt one approach to solve the DSGE model identification problem.
    2. $\Longrightarrow$ Identify only a subset of RBC model parameters after pre-setting or calibrating the remaining parameters.
- Analysis by Hall (1996, "Overtime, effort, and the propagation of business cycle shocks," *JME* 38, 139–160) suggests a reason for this practice.
    1. Whether ML or GMM is being used, these estimators rely on the same (unconditional) sample (means) and theoretical (steady state) information about first moments to identify DSGE model parameters.
    2. Although ML is a full information estimator, which engages all the moment conditions expressed by the DSGE model, GMM and ML rely on the same first moment information for identification.
- The frequentist assumption of a true model binds the identification problem to the issue of DSGE model misspecification.
    1. Can any parameters of a DSGE model be identified when it is misspecified?
    2. For example, frequentist ML can lose its appeal when models are known to be misspecified $\Longrightarrow$ could use quasi-MLE instead.
- Misspecification is a problem about the population properties of a model $\Longrightarrow$ no amount of data or computing power will solve problems related to the identification and misspecification of DSGE models.

Introduction: Identification of DSGE Models, II

- A frequentist response to DSGE identification problems is II.
  1. Smith (1993) and Gourieroux, Monfort, and Renault (1993) develop an II estimator and specification tests with standard asymptotic properties even though the true likelihood of the DSGE model is unknown.
  2. The II approach to DSGE model estimation and evaluation is anticipated by Gregory and Smith (1990, "Calibration as estimation," *Econometric Reviews* 9, 57–89) and (1991, "Calibration as testing: Inference in simulated macroeconomic models," *JBES* 9, 297–303).
  3. The II estimator minimizes a GMM-like criterion in the distance between a vector of theoretical and sample moments, which need to be observed in sample data and predicted by the DSGE model.
  4. Estimating DSGE model parameters is "indirect" because the GMM-like criterion matches sample and theoretical moments not explicitly tied to the DSGE model $\Longrightarrow$ not the likelihood of the DSGE model.
  5. Theoretical moments are produced by simulating synthetic data from the solution of the DSGE model.
  6. Apply a classical optimizer to the GMM-like criterion $\Longrightarrow$ maximize its objective by moving the theoretical moments closer to the sample moments by updating the DSGE model parameters while holding the structural shock innovations fixed.
  7. CEE (2005) estimate a NKDSGE model by matching its theoretical impulse responses to those of an estimated SVAR $\Longrightarrow$ an II estimator.
  8. A discussion of identification problems facing this versiion of II is found in Canova and Sala (2009, "Back to square one: Identification issues in DSGE Models," *JME* 56, 431–449).

Introduction: Identification of DSGE Models, III

- Dridi, Guay, and Renault (2007) extend the II by recognizing DSGE models are false.
  1. A DSGE model is an abstraction of reality and hence misspecified.
  2. Separate the part of a DSGE model having economic content from the misspecified part.
  3. $\Longrightarrow$ Split the vector of DSGE model parameters Γ into the parameters having compelling economic content, Γ_1, and the remaining nuisance or pseudo-parameters, Γ_2, that lack economic interest.
  4. Cannot ignore Γ_2 because it is integral to the DSGE model $\Longrightarrow$ fix Γ_2 or calibrate it with sample information.
  5. Γ_2 contributes to identifying Γ_1, but without polluting it with the misspecification of the DSGE model encapsulated by Γ_2.
  6. $\Longrightarrow$ Dridi, Guay, and Renault construct an asymptotic distribution of Γ_1 that accounts for misspecification of the DSGE model.
  7. The sampling theory is useful for testing misspecification of the DSGE model and to gauge its ability to match the data.

Introduction: Identification of DSGE Models, IV

- Does identification of DSGE models matter for Bayesians?
  1. Many Bayesians argue identification rests on having a well posed prior $\Longrightarrow$ a proper prior independent of the data and has a density that integrates to one.
  2. If the data are uninformative, prior and posterior distributions can be equivalent; see Poirier (1998, "Revising beliefs in nonidentified models," *Econometric Theory* 14, 483–509).

- This problem differs from identification problems frequentists face.
  1. Identification of a model is a problem that arises in population for a frequentist estimator, while
  2. for a Bayesian the source of the equivalence is data interacting with the prior.
  3. Poirier provides analysis suggesting that Γ be split into those parameters for which the data are informative, Γ_1, given the priors from those, Γ_2, for which this is not possible.

Introduction: Identification of DSGE Models, V

- Bayesians avoid having to assume there exists a true or correctly specified DSGE model because of the likelihood principle (LP).
  1. The LP is a foundation of Bayesian statistics $\Longrightarrow$ all evidence about a DSGE model is contained in its likelihood conditional on the data; see Berger and Wolpert (1988, The Likelihood Principle, Second Edition, Hayward, CA: Institute of Mathematical Statistics).
  2. Bayesian likelihood-based evaluation is consistent with the belief "all models are false" $\Longrightarrow$ there is no true DSGE model because this class of models is afflicted with incurable misspecification.

- Bayesian model evaluation tools are also consistent with another widely expressed belief of macroeconomists, "It takes a model to beat a model."
  1. Since the likelihood of a DSGE model summarizes its probabilistic evaluation by the data, the likelihoods of a suite of DSGE models contains all the evidence needed to assess which "best" fits the data.
  2. This suggests Bayesian estimation of DSGE models is not an ends in itself, but a means to the end of conducting model evaluation.

Introduction: The Posterior of DSGE Models, I

- There exist several Bayesian approaches to estimate DSGE models.
- Likelihood-based Bayesian estimation constructs the posterior distribution, $\mathcal{P}(\Gamma \,|\, \mathcal{Y}_{1:T})$, of DSGE model parameters conditional on sample data $\mathcal{Y}_{1:T}$ of length T.
- Bayesian estimation exploits the fact that the posterior distribution equals the DSGE model likelihood, $\mathscr{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma\Big)$, multiplied by the econometrician's priors on the DSGE model parameters, $\mathcal{P}(\Gamma)$, up to a factor of proportionality

$$\mathcal{P}\Big(\Gamma \Big| \mathcal{Y}_{1:T}\Big) \propto \mathscr{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma\Big)\mathcal{P}\Big(\Gamma\Big). \qquad (\textbf{EST}.1)$$

- Bayesian estimation of DSGE models is confronted by posterior distributions too complicated to evaluate analytically.
- The complication arises because the mapping from a DSGE model to its $\mathscr{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma\Big)$ is nonlinear in $\Gamma \Longrightarrow$ engage simulation methods to approximate $\mathcal{P}(\Gamma|\mathcal{Y}_{1:T})$.
- Among the earliest examples of Bayesian likelihood-based estimation of a DSGE model is DeJong, Ingram, and Whiteman (2000a, "A Bayesian approach to dynamic macroeconomics," *JofE* 98, 203–223) and (2000b, "Keynesian impulses versus solow residuals: Identifying sources of business cycle fluctuations," *JAE* 15, 311–329).
  1. Use importance sampling (IS) to compute the posterior as functions of Γ.
  2. A drawback of IS is that it is often unreliable when Γ has large dimension.
  3. Another is there is little guidance about updating $\mathcal{P}(\Gamma \,|\, \mathcal{Y}_{1:T})$ in the IS sampler.

Introduction: The Posterior of DSGE Models, II

- Perhaps, the first instance of Metropolis-Hasting (MH-)MCMC simulation applied to DSGE model estimation is Otrok (2001, "On measuring the welfare cost of business cycles," *JME* 47, 61–92).
- The MH algorithm proposes to update Γ using a multivariate random walk,
  1. but first an initial draw of Γ from $\mathcal{P}(\Gamma)$ is needed.
  2. The initial Γ is updated by adding to it draws from a distribution of "shock innovations" $\Longrightarrow$ the MH algorithm proposes to update Γ using a multivariate random walk.
  3. The decision to keep the initial Γ or to move to the updated Γ depends on whether the latter increases $\mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \Gamma\right)$.
  4. Repeat this procedure by sampling from the proposal distribution generated by the multivariate random walk to update Γ.

Introduction: The Posterior of DSGE Models, III

- Bayesian evaluation of estimated DSGE models relies on the Bayes factor

$$\mathcal{B}_{j,s\,|\,y_{1:T}} = \frac{\mathcal{L}\left(y_{1:T} \,\middle|\, \Gamma_j,\, \mathcal{M}_j\right)}{\mathcal{L}\left(y_{1:T} \,\middle|\, \Gamma_s,\, \mathcal{M}_s\right)}. \qquad \textbf{(EST.2)}$$

  1. The Bayes factor measures the odds the data prefer DSGE model j, $\mathcal{M}_j$, over DSGE model s, $\mathcal{M}_s$, given Γ_j and Γ_s,
  2. which involves the ratio of *marginal* likelihoods of $\mathcal{M}_j$ and $\mathcal{M}_s$.
  3. A marginal likelihood integrates Γ_j out of $\mathcal{L}\left(y_{1:T}\,|\,\Gamma_j,\, \mathcal{M}_j\right)$.
  4. Multiply $\mathcal{B}_{j,s|y_T}$ by the prior odds to find the *posterior odds ratio* $\Longrightarrow \mathcal{R}_{j,s|y_{1:T}} = \mathcal{B}_{j,s|y_{1:T}}\mathcal{P}(\Gamma_j)/\mathcal{P}(\Gamma_s)$.
  5. The log of $\mathcal{R}_{j,s|y_{1:T}}$ is the log of $\mathcal{B}_{j,s|y_{1:T}}$ net of the log of the prior odds of these DSGE models, $\mathcal{P}(\Gamma_j)/\mathcal{P}(\Gamma_s)$.

- The foundations of Bayesian evaluation of DSGE models are covered by
  1. Geweke (1999, "Simulation methods for model criticism and robustness analysis," in Berger, Bernado, Dawid, Smith (eds.), *Bayesian Statistics, Vol. 6*, Oxford, UK: Oxford University Press, 275–299) and (2005, Contemporary Bayesian Econometrics and Statistics, Hoboken, NJ: J. Wiley & Sons, Inc).
  2. Fernández-Villaverde and Rubio-Ramírez (2004, "Comparing dynamic equilibrium models to data: A Bayesian approach," *JofE* 123, 153–187).
  3. The fit of several NKDSGE models are assessed using Bayes factors by Rabanal and Rubio-Ramírez (2005, "Comparing New Keynesian models of the business cycle: A Bayesian approach," *JME* 52, 1151–1166).

Match the Solution to the Linearized NKDSGE Model to the ABCDs Model

- The MH algorithm depends on the likelihood of the linearized NKDSGE model to produce posterior distributions of its parameters.
- Lets give the solution (**SLZ**.7) of the linearized NKDSGE model a state space representation.
- The solution (**SLZ**.7) to the linearized NKDSGE model has the same form as the system of state equations (**KF**.1) of the Kalman filter by equating $\mathcal{S}_t = \mathcal{S}_t$, $\xi_t = \zeta_t$, $\boldsymbol{\mathcal{A}} = \boldsymbol{\Theta}_{\mathcal{S}}$, $\boldsymbol{\mathcal{B}} = \boldsymbol{\Theta}_{\zeta}$, and $\mathbf{E}_t\zeta_{t+j} = 0 \Longrightarrow$

$$\mathcal{S}_{t+1} = \boldsymbol{\mathcal{A}}\mathcal{S}_t + \boldsymbol{\mathcal{B}}\xi_{t+1}, \quad \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}' = \boldsymbol{\mathcal{Q}}, \quad \xi_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad (\textbf{EST}.3)$$

  where standard deviations are found in the diagonal matrix $\boldsymbol{\mathcal{B}}$.

- The system of observation equations are revised to

$$\mathcal{Y}_t = \mathcal{F}\mathcal{Z}_t + \boldsymbol{\mathcal{C}}\mathcal{S}_t + \boldsymbol{\mathcal{D}}\omega_t, \quad \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}' = \boldsymbol{\mathcal{R}}, \quad \omega_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad (\textbf{EST}.4)$$

  1. where $\mathcal{F}$ is a column vector containing parameters aimed at estimating the sample means of $\mathcal{Y}_t \Longrightarrow \mathcal{Z}_t = 1$,
  2. $\boldsymbol{\mathcal{C}}$ maps several of the theoretical constructs in $\mathcal{S}_t$ into the actual data $\mathcal{Y}_t$,
  3. ω_t consists of white noise measurement errors associated with $\mathcal{Y}_t$,
  4. and $\boldsymbol{\mathcal{D}}$ is a diagonal matrix of measurement error standard deviations.

Review: The Kalman Filter Generates the Likelihood of the Linearized NKDSGE Model

- The likelihood (**KS**.7) of the linearized NKDSGE model, which is built up by generating forecasts from the state space system (**EST**.3) and (**EST**.4) period-by-period, is restated as

$$\mathcal{L}\left(y_{1:T} \,\middle|\, \Gamma\right) = \prod_{t=1}^{T} f_{y_t \,|\, y_{t-1}}\left(y_t \,\middle|\, y_{t-1}\right). \qquad (\textbf{EST}.5)$$

  where Γ contains the entire set of deep structural NKDSGE model parameters.

- The following steps compute the likelihood (**EST**.5) using the Kalman filter:
  1. Set $\mathcal{S}_{1|0} = \mathbf{0}$ and calculate $\mathbf{P}_{1|0}$ using (**KF**.4).
  2. Construct $\mathcal{Y}_{1|0} = \mathcal{F} + \mathcal{C}\mathcal{S}_{1|0} = \mathcal{F}$ and employ (**KF**.8) to produce the MSE of $\mathcal{Y}_{1|0}$ $\Longrightarrow \boldsymbol{\Omega}_{y,1|0} = \mathcal{C}\,\mathbf{P}_{1|0}\,\mathcal{C}' + \mathcal{R}$ (implying $\boldsymbol{\Omega}_{y,t|t-1} = \mathcal{C}\,\mathbf{P}_{t|t-1}\,\mathcal{C}' + \mathcal{R}$).
  3. The predictions made in Steps 1 and 2 yield the date 1 likelihood

$$\mathcal{L}\left(y_1 \,\middle|\, \Gamma\right) = (2\pi)^{-0.5n} \left|\boldsymbol{\Omega}^{-1}_{1|0}\right|^{0.5} \exp\left[-\frac{1}{2}\left(y_1 - \mathcal{F}\right)'\boldsymbol{\Omega}^{-1}_{y,1|0}\left(y_1 - \mathcal{F}\right)\right].$$

  4. Next, update $\mathcal{S}_{1|1}$ and its MSE $\mathbf{P}_{1|1}$ using equations (**KF**.10) and (**KF**.11).
  5. Repeat steps 2, 3, and 4 for $t = 2, \ldots, T$ to obtain KF predictions of $\mathcal{S}_{t+1|t}$, $\mathbf{P}_{t+1|t}$, $\mathcal{S}_{t|t}$, and $\mathbf{P}_{t|t}$ engaging equations (**KF**.16), (**KF**.18), (**KF**.15), and (**KF**.17) to produce the forecast $\mathcal{Y}_{t|t-1} = \mathcal{F} + \mathcal{C}\mathcal{S}_{t|t-1}$, its MSE $\boldsymbol{\Omega}_{y,t|t-1}$, and the likelihood,

$$\mathcal{L}\left(y_t \,\middle|\, y_{t-1}, \Gamma\right) = (2\pi)^{-0.5n} \left|\boldsymbol{\Omega}^{-1}_{y,t|t-1}\right|^{0.5} \exp\left[-\frac{1}{2}\left(y_t - \mathcal{Y}_{t|t-1}\right)'\boldsymbol{\Omega}^{-1}_{y,t|t-1}\left(y_t - \mathcal{Y}_{t|t-1}\right)\right].$$

Choosing Priors for the Linearized NKDSGE Model, I

- The likelihoods $\mathcal{L}\big(y_1 \,\big|\, \Gamma\big)$, $\mathcal{L}\big(y_{2:1} \,\big|\, y_1, \Gamma\big)$, $\mathcal{L}\big(y_{3:1} \,\big|\, y_{2:1}, \Gamma\big)$, $\ldots$, $\mathcal{L}\big(y_{t:1} \,\big|\, y_{t-1:1}, \Gamma\big)$, $\ldots$, $\mathcal{L}\big(y_{T-1:1} \,\big|\, y_{T-2:1}, \Gamma\big)$, and $\mathcal{L}\big(y_{T:1} \,\big|\, y_{T-1:1}, \Gamma\big)$ computed at Steps 2 and 5 are used to build up the likelihood function (**EST**.5) of the linearized NKDSGE model.

- Construction of the posterior of Γ, $\mathcal{P}(\Gamma \,|\, y_{1:T})$, needs these likelihoods and the prior of Γ, $\mathcal{P}(\Gamma)$; see (**EST**.1).

- Del Negro and Schorfheide (2008) discuss methods to construct $\mathcal{P}(\Gamma)$.
  1. First, break the NKDSGE model parameters into three sets; see table 2 of Del Negro and Schorfheide (2008, p. 1201).
  2. The first set contains parameters that define the steady state of the NKDSGE model $\Longrightarrow$ Hall (1996) shows ties the steady state of the NKDSGE model to the unconditional first moments of $y_{1:T}$.
  3. The second set of parameters rely on preferences, technologies, and market structure that affect mechanisms endogenously propagating exogenous shocks.
  4. Along with technology, preference, and market structure parameters, Del Negro and Schorfheide add parameters of the Taylor rule (**NK**.9) to this set; see the agnostic sticky price and wage priors of tables 1 and 2 of Del Negro and Schorfheide (2008, pp. 1200–1201).
  5. The third set of parameters have AR1 coefficients and standard deviations of the exogenous shocks; see table 3 of Del Negro and Schorfheide (2008, p. 1201).

Choosing Priors for the Linearized NKDSGE Model, II

- Separate the parameter vector Γ into two parts, where the parameters of economic interest are placed in 18×1 column vector

$$\Gamma_1 = \left[h \ \gamma \ \alpha \ \psi \ N_0 \ \delta \ \xi \ \mu_p \ \theta \ \mu_W \ a'' \ \varpi \ \kappa_\pi \ \kappa_y \ m^* \ \rho_R \ \sigma_\varepsilon \ \sigma_\upsilon \right]'.$$

- Next group the relevant elements of Θ_1 into the steady state parameter vector

$$\Gamma_{1,\text{ss}} = \left[\alpha \ N_0 \ \xi \ \theta \ m^* \right]'.$$

- The parameters responsible for endogenous propagation in the NKDSGE model

$$\Gamma_{1,\text{prop}} = \left[h \ \gamma \ a'' \ \mu_p \ \mu_W \ \kappa_\pi \ \kappa_y \ \rho_R \right]'.$$

- The source of fluctuations in the NKDSGE model are the innovations to the exogenous shocks with standard deviations

$$\Gamma_{1,\text{exog}} = \left[\sigma_\varepsilon \ \sigma_\upsilon \right]'.$$

- If the Taylor rule (**NK**.9) is replaced by the money supply growth rule (**NK**.8), its parameters, ρ_m and σ_μ, are added to $\Gamma_{1,\text{exog}}$ and σ_υ is deleted as is ρ_R from $\Gamma_{1,\text{prop}}$.

Choosing Priors for the Linearized NKDSGE Model, III

- Often priors for $\Gamma_{1,ss}$, $\Gamma_{1,prop}$, and $\Gamma_{1,exog}$ are drawn from normal, beta, gamma, and inverse gamma distributions, but not restricted to these distributions; details are in Del Negro and Schorfheide (2008).
    1. Priors reflect your uncertainty about the NKDSGE model, but employ priors that are easy to communicate to the audience.
    2. Report summary statistics of your priors $\Longrightarrow$ a normal prior consists of its mean and standard deviation while other distributions are define by one or more parameters (*i.e.*, a gamma or inverse gamma distribution have scale and shape parameters).
    3. As examples endow the consumption habit parameter $h \sim$ beta with mean = 0.65 and standard deviation = 0.10 $\Longrightarrow$ the 95% coverage interval is $[0.22, 0.96]$,
    4. $\gamma \sim$ gamma with a shape parameter = 2.0 and a scale parameter = 0.750 $\Longrightarrow$ the 95% coverage interval is $[0.25, 3.57]$,
    5. $\rho_R \sim$ normal with mean 0.50 and standard deviation 0.03 $\Longrightarrow$ the 95% coverage interval is $[0.26, 0.74]$, and
    6. $\sigma_\varepsilon \sim$ inverse gamma with a shape parameter = 4.0 and a scale parameter = 0.3 $\Longrightarrow$ the 95% coverage interval is $[0.03, 0.28]$.
    7. Present 95% coverage intervals of the priors to give a sense of your uncertainty of the NKDSGE model parameter by parameter.

Choosing Priors for the Linearized NKDSGE Model, IV

- The remaining parameters fall into

$$\Gamma_2 = \begin{bmatrix} \beta & \psi & \delta & \mathcal{L}_{AF} \end{bmatrix}'.$$

- Priors are not imposed on the steady state nominal interest rate R^* and the steady state ratios, C^*/Y^*, X^*/K^*, and K^*/Y^*.
  1. Instead, calibrate β (possibly using information about the sample mean of $R_t - \pi_{t+1}$) to estimate m^* within the MH algorithm.
  2. Next, fix the capital share, α and the depreciation rate, δ to ease the burden of identifying elements in $\Gamma_{1,ss}$
  3. This approach to fixing or calibrating several model parameters dates to the first attempts at estimating RBC models.

- Often the steady state of a NKDSGE model implies theoretical means of the elements of $\mathcal{Y}_t$ that are far from the associated sample mean.
  1. Del Negro and Schorfheide (2008, p. 1197) suggest including a constant or "add-factor" in one or more of the measurement equations (**EST**.4) to bridge the gap between the theoretical and sample means.
  2. This is to equivalent to adding $\mathcal{L}_{AF}$ to the log likelihood, $\ln\mathcal{L}\left(\mathcal{Y}_{1:T} \,\middle|\, \Gamma_1, \Gamma_2\right)$, of the linearized NKDSGE model as $\ln\mathcal{L}\left(\mathcal{Y}_{1:T} \,\middle|\, \Gamma_1, \Gamma_2\right) + \ln\mathcal{L}_{AF}$.
  3. $\Longrightarrow$ Incorporate "extra" information into the log likelihood to prevent inconsistent estimates of the innovations $\mathcal{Y}_t - \mathcal{Y}_{t|t-1}$.

A Short Aside on Markov Chain Monte Carlo Methods, I

- Estimation of linearized NKDSGE models is hindered by the problem of mapping
  1. from $\mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \Gamma_1, \Gamma_2\right)\mathcal{P}\left(\Gamma_1\right)$ to $\mathcal{P}\left(\Gamma_1 \middle| \mathcal{Y}_T, \Gamma_2\right)$ inherent in (**EST**.1) is non-analytic.
  2. $\Longrightarrow$ The posterior $\mathcal{P}\left(\Gamma_1 \middle| \mathcal{Y}_T, \Gamma_2\right)$ has a non-standard distribution.

- Classical simulation methods (*i.e.*, II) or importance sampling (IS) could be applied to estimate NKDSGE model, but a high dimensional Γ_1 creates problems.
  1. Although the KF makes is easy to construct $\mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \Gamma_1, \Gamma_2\right)$, its numerical optimization with respect to Γ_1 with many elements is difficult.
  2. There is little guidance about choosing an importance distribution, $\mathcal{G}(\Gamma_1)$, that reliably mimics $\mathcal{P}\left(\Gamma_1 \middle| \mathcal{Y}_T, \Gamma_2\right)$, especially across repeated draws from $\mathcal{G}(\Gamma_1)$.

- Markov Chain Monte Carlo (MCMC) methods provide relief to the problem of computing $\mathcal{P}\left(\Gamma_1 \middle| \mathcal{Y}_T, \Gamma_2\right)$ when Γ_1 is of high dimension.
  1. A Markov chain is a stochastic process, which evolves according to the transition kernel $p(z, y) \Longrightarrow$ the conditional density of y given z or $p(y|z)$.
  2. The kernel $p(z, y)$ is the continuous analogue to the discrete transition probability $p_{i,j} = P(Z_{t+1} = j | Z_t = i)$, which is probability of moving from state i to state j

A Short Aside on Markov Chain Monte Carlo Methods, II

- A Markov chain is so named because it has the Markov property

$$f_{(Z_1,\ldots,Z_T \,|\, Z_0 = z_0)}(z_1, \ldots, z_T) = p(z_0, z_1)p(z_1, z_2)\ldots p(z_{T-2}, z_{T-1})p(z_{T-1}, z_T).$$

  1. Given $Z_0 = z_0$, assume the joint density of $\{Z_t\}_{t=1}^{T}$ is the multiplicative sum of the kernels (*i.e.*, the conditional densities) of the realized states $\{z_t\}_{t=1}^{T}$.
  2. $\Longrightarrow$ Conditional on the current state z, the probability the stochastic process realizes a value in S, where S is subset of the real line, $S \subseteq R$, is denoted

$$P(z, S) = \int_S p(z, y)dy, \text{ and the } j\text{th step ahead probability is } P^j(z, S) = \int_R P(z, dy)P^{j-1}(y, S).$$

- Next, define an *invariant* density π_y for the kernel $p(z, y)$

$$\pi_y = \int_R \pi_z p(z, y)dz.$$

  1. $\Longrightarrow$ The probability the stochastic process is in state y at any date t is marginalized (*i.e.*, remove or integrate out all other states) with respect to the state at date $t-1$.
  2. Or π_y answers the question, "What is the probability of being in state z at date $t-1$, π_z, and transiting to state y at date t with probability $p(z, y)$?"
  3. The (unconditional) probabilities are restricted by $\int \pi_y dy = 1$.

A Short Aside on Markov Chain Monte Carlo Methods, III

- Is the invariant density π_y unique presuming it exists?
- **Definition**: A Markov chain is *irreducible* if a (non-negligible) positive probability exists of moving from any state z at date t to any state y in a finite span of time.
- **Definition**: A Markov chain is *recurrent* if every state z has probability $= 1$ of being "hit" (at least once) in a finite span of time.
- **Definition**: A Markov chain is *aperiodic* if the number of steps needed to move from any z to itself is not a constant integer (or a multiple thereof), where the existence of only one aperiodic state is necessary to show the Markov chain is irreducible.
- **Theorem**: Assume $\pi.$ is an invariant distribution for $P(z,\,\cdot)$ and it is $\pi.$-irreducible $\Longrightarrow P(\cdot,\,\cdot)$ is positive recurrent (exists) and its unique invariant distribution is $\pi.$.
  1. The idea is that $\lim_{j\longrightarrow\infty} P^j(\cdot,\,\cdot)$ converges to $\pi.$ for almost all z.
  2. This holds for all z if the number of times this state is realized is unbounded $\Longrightarrow P(z,\,\cdot)$ is Harris recurrent.
- **Theorem**: Given $P(z,\,\cdot)$ is positive recurrent (exists) and its unique invariant distribution is $\pi.$, it is $\pi.$-irreducible aperiodic, positive recurrent, and its invariant distribution is $\pi.$.

Why Do MCMC Algorithms Differ from Standard Monte Carlo Methods?

- Consider a univariate random variable Z_t drawn from a (stationary) *IID* distribution.
- Suppose there is function of Z_t, $k(Z_t)$, which is a known stochastic process, but is not analytic, where the goal is to estimate the first two moments of Z_t.
  1. Denote the first population moment as $\mu_Z = \mathbf{E}\{k(Z_t)\}$, but cannot be estimated using sample data $\Longrightarrow$ use a central limit theorem to show the estimator of μ_Z is $\widehat{\mu}_Z \sim \mathcal{N}\left(\mu_Z,\, T^{-1}\sigma_Z^2\right)$, where the second population moment is $\sigma_Z^2 = \text{var}(Z_t)$.
  2. Assume there is way to simulate $\{Z_t\}_{t=1}^{T} \Longrightarrow$ using Monte Carlo methods approximate $\widehat{\mu}_Z = \sum_{t=1}^{T} k(z_t)$, where z_t are simulated draws from $Z_t \sim$ *IID*.
  3. The simulation estimator of the variance of Z_t is $\widehat{\sigma}_Z^2 = T^{-1}\sum_{t=1}^{T}\left[k(z_t) - \widehat{\mu}_Z\right]^2$.
- Suppose $\{Z_t\}_{t=1}^{T}$ evolves as a stationary Markov chain $\Longrightarrow$ the conditional distribution of Z_{T+1} relies only on Z_T, not $\{Z_t\}_{t=1}^{T}$, and the unconditional distribution remains *IID*.
  1. The Markovian property does not necessarily carry over to $k(Z_t) \Longrightarrow$ the serial dependence inherent in the Markov chain alters the estimator of σ_Z^2.
  2. The distribution of $\widehat{\mu}_Z$ is unchanged, but population variance of Z_t becomes $\sigma_Z^2 = \text{var}(Z_t) + 2\sum_{j=1}^{\infty}\text{cov}\Big(k(z_t),\, k(z_{t+j})\Big)$.
  3. In practice, have to confront a MCMC simulator inducing serial dependence in $\sigma_Z^2 \Longrightarrow$ need to control this dependence for the simulator to converge.
- Under Harris recurrence, the CLT is valid for any initial distribution and a law of motion for the transition probabilities, given the CLT holds for one initial distribution and the same law of motion $\Longrightarrow$ the MCMC converges to a invariant distribution.

A Short Aside on the Metropolis-Hasting Algorithm, I

- The MC definitions and theorems form the bedrock of MH-MCMC.
- **Definition**: A kernel $q(\cdot, \cdot)$ is reversible, given $f(z)q(z, y) = f(y)q(y, z)$.
- Suppose $f(z)q(z, y) > f(y)q(y, z)$, which is not reversible kernel $\Longrightarrow$ the case for a proposal density in a MH algorithm.
- The Metropolis-Hasting algorithm flips $q(y, z)$ into being a reversible kernel.
  1. Choose a kernel $c(\cdot, \cdot)$ such that $f(z)c(z, y)q(z, y) = f(y)c(y, z)q(y, z)$; see Chib and Greenberg (1995, "Understanding the Metropolis-Hasting algorithm," *The American Statistician* 49, 327–335).
  2. $\Longrightarrow$ The kernel $c(z, y)$ contains information whether there is a high probability in moving to state z from state y given the proposed move is described by the transition kernel is $q(z, y)$.
  3. Given the proposed move, $c(y, z) = 1 \Longrightarrow f(z)c(z, y)q(z, y) = f(y)q(y, z)$, which yields

$$c(z, y) \;=\; \min\left\{\frac{f(y)q(y, z)}{f(z)q(z, y)},\, 1\right\}, \;\text{ or }\; c(z, y) \;=\; 0.$$

  4. The MH density $c(z, y)$ serves as the criterion to decide whether to accept the move to state z from state y.
  5. Note the restriction $f(z)q(z, y) \neq 0$ is necessary for $c(z, y)$ to exist, but, as we will see, for constructing $\mathcal{P}\left(\Gamma_1 \,\middle|\, \mathcal{Y}_T, \Gamma_2\right)$ this should not be an issue.

A Short Aside on the Metropolis-Hasting Algorithm, II

- How to choose $q(\cdot, \cdot)$ to operate the MH-MCMC?
- MH-MCMC aims to approximate the posterior, say, of a parameter vector y by simulation.
  1. The Metropolis part of MH algorithm restricts innovations to the proposal distribution be drawn from a symmetric distribution. (The Hastings part loosens this restriction on the innovations to the proposal updating law of motion.)
  2. The MH law of motion proposes an update of the posterior starting from the most recent acceptance to the posterior plus the symmetrically distributed innovations $\Longrightarrow$ equate $q(\cdot, \cdot)$ with a random walk plus Gaussian innovations.
  3. Thus, $q(y, z) = q(z, y) \Longrightarrow$ innovations to the random walk are drawn from a symmetric distribution.
  4. An implication is a move to a higher point in the density y from z, $f(y) > f(z)$, occurs with certainty while the converse can occur but with probability < 1.
  5. The result is the MH decision criterion becomes

$$c(z, y) \; = \; \min\left\{\frac{f(y)}{f(z)}, \, 1\right\}.$$

A Short Aside on the Metropolis-Hasting Algorithm, III

- What is $f(\cdot)$?

- The MH algorithm draws from a proposal distribution with transition law of motion $q(y, z)$ to construct the posterior $\mathcal{P}(y)$.
  1. The only restriction is proportionality of $f(y)$ and $\mathcal{P}(y) \Longrightarrow \mathcal{P}(y) \propto f(y)$.
  2. This is a weak restriction, which helps to explain the application of the MH algorithm to many different kinds of estimation problems.
  3. $\Longrightarrow$ No need to find a $f(y)$ that is an exact replication of $\mathcal{P}(y)$.
  4. The MH algorithm dispenses with the need to compute the normalizing constant of $f(y)$.
  5. $\Longrightarrow$ In this case, the normalizing constant is the divisor of $f(y)$ that forces its probability density function to sum to one.

- The weak proportionality restriction means the MH algorithm is a tool
  1. to obtain the posterior of a model by equating $f(\cdot)$ to the likelihood of the model crossed by the prior distribution of the model coefficients.
  2. $\Longrightarrow$ This proportionality relationship is described by (**EST**.1) for the NKDSGE model and is reproduced here

$$\mathcal{P}\Big(\Gamma_1 \Big| \mathcal{Y}_{1:T}, \Gamma_2\Big) \propto \mathcal{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma_1, \Gamma_2\Big)\mathcal{P}\Big(\Gamma_1\Big).$$

Overview of MH-MCMC Sampling for the NKDSGE Model

- The posterior distribution of the NKDSGE model parameters in Γ_1 is characterized using the MH-MCMC algorithm.
  1. The MH-MCMC algorithm is initialized with Γ_1 conditional on $\Gamma_2 \Longrightarrow$ need a "guess" of Γ_1 and calibration of Γ_2 to begin the sampler.
  2. The initial Γ_1 is passed to the Kalman filter routines described by equations (**KF**.15)–(**KF**.18) to produce an *estimate* of $\ln \mathcal{L}\Big(\mathcal{Y}_{1:T}\,\Big|\,\Gamma_1,\,\Gamma_2\Big)$.
  3. Next, employ the MH random walk law of motion to update the initial Γ_1 to a new proposal of the posterior of the linear approximate NKDSGE model.
  4. Input the proposed update of Γ_1 into the Kalman filter to generate a second estimate of the likelihood of the linear approximate NKDSGE model.
  5. Engage the MH decision criterion to decide whether the initial or proposed update of Γ_1 and the associated likelihood passes into the next step of sampler.
  6. Given this decision, the next step of the MH algorithm obtains a new proposal of Γ_1 using the random walk law of motion, which produces an estimate of the likelihood at this new update.
  7. Compare the updated likelihood with the likelihood obtained in the previous MH step using the MH decision criterion to choose the likelihood and Γ_1 carried into the next MH step.
  8. Repeat the process $\mathcal{J}$ steps to create the posterior of the linear approximate NKDSGE model, $\mathcal{P}(\Gamma_1\,|\,\mathcal{Y}_{1:T};\,\Gamma_2)$.

A MH-MCMC Algorithm, I

- Apply the MH-MCMC sampler to the linearized NKDSGE model in the following steps.
  1. Initialize the MH algorithm with the vector of NKDSGE model parameters $\widehat{\Gamma}_{1,0}$.
  2. Pass $\widehat{\Gamma}_{1,0}$ to the KF routines (**KF**.15)–(**KF**.18) to obtain an initial estimate of the $\mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \widehat{\Gamma}_{1,0}, \Gamma_2\right)$.
  3. Propose an update of $\widehat{\Gamma}_{1,0}$, $\Gamma_{1,1}$, using the MH random walk law of motion

$$\Gamma_{1,1} \;=\; \widehat{\Gamma}_{1,0} \;+\; \vartheta \boldsymbol{\Omega}_{\Gamma_1}^{0.5} \varsigma_1, \quad \varsigma_1 \sim \mathcal{NID}\left(\mathbf{0}_d, \mathbf{I}_d\right),$$

     where ϑ is a scalar that controls the size of the "jump" of the proposed MH random walk update, $\boldsymbol{\Omega}_{\Gamma_1}$ is the Cholesky decomposition of the covariance matrix of Γ_1, and d (= 18) is the dimension of Γ_1. Obtain $\ln \mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \Gamma_{1,1}, \Gamma_2\right)$ by operating the Kalman filter inputting $\Gamma_{1,1}$.
  4. The MH algorithm employs a two-stage procedure to decide whether to keep the initial $\widehat{\Gamma}_{1,0}$ or move to the updated proposal $\Gamma_{1,1}$. First, calculate

$$c_1 \;=\; \min\left\{ \frac{\ln \mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \Gamma_{1,1}, \Gamma_2\right) \mathcal{P}\left(\Gamma_{1,1}\right)}{\ln \mathcal{L}\left(\mathcal{Y}_{1:T} \middle| \Gamma_{1,0}, \Gamma_2\right) \mathcal{P}\left(\widehat{\Gamma}_{1,0}\right)}, \; 1 \right\},$$

     where $\mathcal{P}(\Gamma_{1,1})$ and $\mathcal{P}(\Gamma_{1,0})$ are the priors at $\Gamma_{1,1}$ and $\Gamma_{1,0}$. Second, draw a uniform random variable $u_1 \sim U(0, 1)$ to select $\widehat{\Gamma}_{1,1} = \Gamma_{1,1}$ and the counter $\wp = 1$ if $u_1 \leq c_1$, otherwise $\widehat{\Gamma}_{1,1} = \widehat{\Gamma}_{1,0}$ and $\wp = 0$.

A MH-MCMC Algorithm, II

5. Repeat steps 3 and 4 for $j = 2, 3, \ldots, \mathcal{J}$ using the MH random walk law of motion

$$\Gamma_{1,j} = \widehat{\Gamma}_{1,j-1} + \vartheta \boldsymbol{\Omega}_{\Gamma_1}^{0.5} \varsigma_j, \quad \varsigma_j \sim \mathcal{NID}\Big(\mathbf{0}_{d\times 1}, \mathbf{I}_d\Big), \qquad (\mathbf{EST}.6)$$

and draw the uniform random variable $u_j \sim U(0, 1)$ to test against

$$c_j = \min\left\{ \frac{\ln \mathscr{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma_{1,j}, \Gamma_2\Big) \mathcal{P}\Big(\Gamma_{1,j}\Big)}{\ln \mathscr{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma_{1,j-1}, \Gamma_2\Big) \mathcal{P}\Big(\widehat{\Gamma}_{1,j-1}\Big)}, 1 \right\},$$

to equate $\widehat{\Gamma}_{1,j}$ to $\Gamma_{1,j}$ conditional on $u_j \leq c_j$; otherwise $\widehat{\Gamma}_{1,j} = \widehat{\Gamma}_{1,j-1}$. The latter implies updating the counter to $\wp = \wp + 0$, while the former has $\wp = \wp + 1$.

- Steps 1–5 of the MH-MCMC algorithm produce the posterior as shown in (**EST**.1), $\mathcal{P}(\widehat{\Gamma}_1 \,|\, \mathcal{Y}_{1:T}; \Gamma_2)$, of the linear approximate NKDSGE model by drawing from $\Big\{\widehat{\Gamma}_{1,j}\Big\}_{j=1}^{\mathcal{J}}$.

- The decision in steps 4 and 5 to accept the updated proposal, $\Gamma_{1,j}$, which is grounded on the acceptance criterion $u_j \leq c_j$, is akin to moving to a higher point on the likelihood surface, $\mathscr{L}\Big(\mathcal{Y}_{1:T} \Big| \Gamma_1, \Gamma_2\Big)$.

A MH-MCMC Algorithm, II

- There are several more issues that have to be resolved to run the MH-MCMC algorithm to create $\mathcal{P}(\widehat{\Gamma}_1 | \mathcal{Y}_{1:T}; \Gamma_2)$.

- The issues concern
  1. obtaining an $\widehat{\Gamma}_{1,0}$ to initialize the MH-MCMC,
  2. computing the covariance matrix of Γ_1, $\boldsymbol{\Omega}_{\Gamma_1}$,
  3. determining the number of steps, $\mathcal{J}$, to run the MH-MCMC sampler,
  4. fixing ϑ to achieve the optimal acceptance rate $\wp/\mathcal{J}$ for the proposal $\Gamma_{1,j}$,
  5. and determining whether the MH-MCMC simulator has converged.

- Rules for improving the MH-MCMC simulator
  1. by increasing the efficiency of the random walk law of motion (**EST**.6) for updating the proposal $\Gamma_{1,j}$
  2. to find optimal acceptance rates of the MH-MCMC sampler
  3. are reviewed in Gelman, Carlin, Stern, and Rubin (2004, Bayesian Data Analysis, second edition Chapman and Hall/CRC: Boca Raton, FL) an especially pp. 305–307.

A MH-MCMC Algorithm, III

- Step 1 of the MH-MCMC algorithm fails to explain the process engaged to obtain $\widehat{\Gamma}_{1,0}$.
- Calculate $\widehat{\Gamma}_{1,0}$ using classical optimization methods.
  1. First, draw 100 samples of Γ_1 from $\mathcal{P}(\Gamma_1) \Longrightarrow \Gamma_{1,i}$, $i = 1, \dots, 100$.
  2. Use $\Gamma_{1,i}$ as a starting value to "estimate" Γ_1 by applying a classical optimizer to the log likelihood of the linear approximate NKDSGE model conditional on Γ_2 $\Longrightarrow$ repeat this process for each of the 100 $\Gamma_{1,i}$s.
  3. Use these MLEs to find the mode of the posterior distribution of Γ_1 $\Longrightarrow$ equate the mode of Γ_1 to the initial condition in a "burn-in" stage of the MH-MCMC algorithm.
  4. Or compute this initial condition as the mean or median of the 100 MLE of Γ_1.
- The burn-in stage of the MH-MCMC algorithm aims to eliminate dependence of $\mathcal{P}(\widehat{\Gamma}_1 | \mathcal{Y}_T; \Gamma_2)$ on the initial condition $\widehat{\Gamma}_{1,0}$.
  1. The dependence is annihilated by drawing $\widehat{\Gamma}_{1,0}$ from a distribution that resembles $\mathcal{P}(\widehat{\Gamma}_1 \,|\, \mathcal{Y}_{1:T}; \Gamma_2)$.
  2. Next, $\mathcal{J}_{BR}$ MH steps are run with $\vartheta = 1$ and $\boldsymbol{\Omega}_{\Gamma_1} = \mathbf{I}_d$ to finish the burn-in stage.
  3. The last MH step of the burn-in yields $\widehat{\Gamma}_{1,0}$, which initializes the final $\mathcal{J}_P$ steps of the final stage of the MH-MCMC algorithm, where $\mathcal{J} = \mathcal{J}_{BR} + \mathcal{J}_P$.
  4. Produce $\mathcal{J}_{BR}$ estimates of Γ_1 during the MH burn-in steps to compute an empirical estimate of the covariance matrix $\boldsymbol{\Omega}_{\Gamma_1}$ the MH law of motion (**EST**.6) needs to generate the proposed updates $\Gamma_{1,j}$.

A MH-MCMC Algorithm, IV

- The scale of the "jump" from $\Gamma_{1,j-1}$ to $\widehat{\Gamma}_{1,j}$ determines the speed at which the proposals $\Gamma_{1,j}$ converge to $\mathcal{P}(\widehat{\Gamma}_1 \,|\, \mathcal{Y}_{1:T}; \Theta_2)$ within the MH-MCMC simulator.

- The speed of convergence depends on $\mathcal{J}$ as well as on ϑ and $\boldsymbol{\Omega}_{\Gamma_1}$.
  1. The number of steps $\mathcal{J}_P$ of the final stage of the MH-MCMC simulator has to be sufficient to allow for convergence.
  2. The choice of $\mathcal{J}$ is sensitive to the dimension of Γ_1 and the structure of the NKDSGE model.
  3. Setting $\mathcal{J} = 500{,}000$, where $\mathcal{J}_P = 2\mathcal{J}_{BR}$, is common for NKDSGE models similar to the model discussed in these notes.
  4. Nonetheless, the choice $\mathcal{J}$ is most likely going to many times larger for larger, more complex, and more nonlinear NKDSGE models.

- Another key to control the speed of convergence of the MH-MCMC sampler is the scalar jump coefficient ϑ.
  1. Gelman et al (2004) recommend greatest efficiency of the MH law of motion (**EST**.6) is using $\vartheta = 2.4/\sqrt{d}$.
  2. A practice widely employed when estimating NKDSE models using the MH-MCMC algorithm sampler is to fix ϑ by repeatedly running the algorithm on the linearized NKDSGE model until an acceptance rate $\wp/\mathcal{J} \in [0.25, 0.35]$ is obtained.

A MH-MCMC Algorithm, V

- Standard practice is to test the convergence of the MH-MCMC simulator along with requiring $\wp/\mathcal{J} \in [0.25,\, 0.35]$.

- One check on the convergence of the MH-MCMC simulator is the $\widehat{\mathcal{R}}$ statistic of Gelman et al. (2004).

  1. The $\widehat{\mathcal{R}}$ statistic compares the variances of the elements within the sequence of $\left\{\widehat{\Gamma}_{1,j}\right\}_{j=1}^{\mathcal{J}}$ to the variance across several sequences produced by the MH-MCMC simulator given different initial conditions; see pp. 269–297 of Gelman et al. (2004) for details.
  2. These different initial conditions are produced using the same methods already described with one exception.
  3. The initial condition for the burn-in stage of the MH-MCMC algorithm is typically set at the next largest mode of the posterior distribution obtained by applying the classical optimizer to the likelihood of the linear approximate NKSDSGE model, where this process is repeated, at least, three to five times.
  4. Gelman et al. (2004) suggest the check is $\widehat{\mathcal{R}} < 1.1$ element by element in $\widehat{\Gamma}_1$.
  5. If the check fails for any element of $\widehat{\Gamma}_1$ across the posteriors of the MH-MCMC chains, there is excessive variation relative to the variance within the sequences.
  6. When $\widehat{\mathcal{R}} \gg 1.1$ is large across the d elements of $\widehat{\Gamma}_1$, Gelman et al suggest raising $\mathcal{J}$ until convergence is achieved, according to the check $\widehat{\mathcal{R}} < 1.1$.
  7. Geweke (2005) advocates a convergence test examining the serial correlation within the sequence of each element of $\widehat{\Gamma}_{1,j}$, $j = 1, \ldots, \mathcal{J}$.