– Lecture Notes 1

Review of fundamental concepts in time series analysis

Please let me know if you have any comments or if you find typos.

These notes contains a rough review of linear time series analysis. The techniques reviewed here will be useful for various topics discussing during the length of the course, inclusing the analysis of structural vector auoregressions (SVAR) and the solution and estimation of linearized dynamic stochastic general equilibrium models (DSGE). Here we cover the following topics:

a) Definitions and basic building blocks of linear time series.

b) Linear least squares and recursive projections.

c) Wold representation theorem.

d) Brief review of limit theorems.

Basic time series concepts

A time series is a set of repeated observations of a variable (say, GDP) over a number of periods $t = 1, 2, .., T$. We denote the time series by $\{x_1, x_2, ..., x_T\}$ (or $\{x_t\}_{t=1}^{T}$) and think of it as the realized value of (a chunk of) a stochastic process.

A stochastic process is a collection of random variables indexed by a number t in some set $\mathcal{T}$. We take $\mathcal{T}$ to be the set of integer numbers and think of each $t \in \mathcal{T}$ as a "time period". A time period could be one year, one quarter, one month, and so on. A stochastic process X_t is a collection of random variables

$$\boldsymbol{X} = \{X_t\}_{t=-\infty}^{\infty} = \{...X_{-2}, X_{-1}, X_0, X_1, X_2, ...\}. \tag{1.1}$$

This process extends infinitely into the past and the future. With each drawing of the stochastic process, we draw an entire *sequence* $\{x_t\}_{t=-\infty}^{\infty} \in \boldsymbol{X}$. Our interest lies in studying probability distributions over such sequences. By conceptualizing the stochastic process in this manner, we can use Hilbert space theory's tools to develop formal arguments and proofs, although we will mostly employ an informal approach.

To use Hilbert space theory, it is necessary to impose certain restrictions on the set of random variables under consideration. We assume that each X_t satisfies the condition

$$E\left[X_t^2\right] < \infty. \tag{1.2}$$

Random variables satisfying this condition are referred to as belonging to L^2, which represents the set of all random variables with finite second moments.

Hilbert spaces are the natural generalization of Euclidean spaces into infinite dimensions. In formal terms, a Hilbert space is a complete normed linear space in which the norm is defined through an *inner product*. For a given pair of elements X, Y belonging in L^2, we define the inner product as

$$\langle X, Y\rangle = E\left[XY\right].$$

The norm associated with this inner product is given by:

$$\|X\| = \langle X, X\rangle^{1/2} = \left(E\left[X^2\right]\right)^{1/2}.$$

We now prove a classic lemma that is widely used.

Lemma (Cauchy-Schwarz inequality)**.** *Let $X, X \in L^2$. Then,*

$$\left|E\left[xy\right]\right| \leq E\left[x^2\right]^{1/2} E\left[y^2\right]^{1/2}. \tag{1.3}$$

Proof. Note that, for any realization of X and Y,

$$\begin{aligned} 0 \leq & \left(\frac{|X|}{E\left[X^2\right]^{1/2}} - \frac{|Y|}{E\left[Y^2\right]^{1/2}}\right)^2 \\ = & \frac{|X|^2}{E\left[X^2\right]} + \frac{|Y|^2}{E\left[Y^2\right]} - 2\frac{|X|\,|Y|}{E\left[x^2\right]^{1/2} E\left[Y^2\right]^{1/2}}. \end{aligned}$$

Rearranging gives

$$\frac{|X|\,|Y|}{E\left[X^2\right]^{1/2} E\left[Y^2\right]^{1/2}} \leq \frac{1}{2}\left[\frac{|X|^2}{E\left[X^2\right]} + \frac{|Y|^2}{E\left[Y^2\right]}\right].$$

Taking expectations on both sides of the inequality and multiplying by $E\left[X^2\right]^{1/2} E\left[Y^2\right]^{1/2}$ gives $E\left|X\right|\left|Y\right| \leq E\left[X^2\right]^{1/2} E\left[Y^2\right]^{1/2}$. But $\left|E\left[XY\right]\right| \leq E\left[\left|X\right|\left|Y\right|\right]$, which leads to $\left|E\left[XY\right]\right| \leq E\left[X^2\right]^{1/2} E\left[Y^2\right]^{1/2}$. ■

In linear time series analysis, we usually characterize the probability distribution of a stochastic process using means and covariances. With normal distributions, this is all we need to characterize the probability distribution. We denote the (unconditional) mean and covariances of the process $\{X_t\}$ by

$$\begin{aligned} \mu_t = & E\left[X_t\right] \\ \sigma_{t,s} = & E\left[\left(X_t - \mu_t\right)\left(X_s - \mu_s\right)\right]. \end{aligned}$$

We say that a stochastic process in L^2 is *covariance stationary* if μ_t is the same for all t (that is, $E[x_t] = \mu$), and the covariance between X_t and X_s depends only on $t-s$, that is, the distance between time periods and not the particular calendar dates t and s. Other names for the same property—which we use interchangeably—are *weakly stationary*, *wide sense stationary*, *second order stationary*, or simply *stationary*.

The *covariogram* or *autocovariance function* is the sequence of covariances

$$\gamma(\tau) \equiv \sigma_{t,t-\tau} = E[(X_t - \mu)(X_{t-\tau} - \mu)], \tag{1.4}$$

where we already assumed stationarity (otherwise we should use $\gamma_t(\tau)$). The covariogram is symmetric, $\gamma(\tau) = \gamma(-\tau)$.[1]

Moreover, the Cauchy-Schwarz inequality implies

$$|E[(X_t - \mu)(X_{t-\tau} - \mu)]| \leq E\left[(X_t - \mu)^2\right]^{1/2} E\left[(X_{t-\tau} - \mu)^2\right]^{1/2}$$

or $|\gamma(\tau)| \leq \gamma(0)$ for all τ.

Because we can always extract the mean from any stationary time series, it saves a lot on notation to work with mean zero stochastic processes. If we want to recover the mean, we simply add it back to the process. So, from now on, we will mostly consider stochastic processes with zero mean.

A usual way to attach a probability distribution to X_t is through linear combinations of a serially uncorrelated stochastic process ε_t satisfying

$$\begin{aligned} E[\varepsilon_t] &= 0 \text{ for all } t, \\ E\left[\varepsilon_t^2\right] &= \sigma^2 \text{ for all } t, \\ E[\varepsilon_t \varepsilon_{t-\tau}] &= 0 \text{ for all } t \text{ and } \tau \neq 0. \end{aligned}$$

The process ε_t is covariance stationary and is referred to as a **white noise**.

Consider the random variable X_t defined as

$$x_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \tag{1.5}$$

where $\{\varepsilon_t\}$ are white noise shocks, and the sequence $\{a_j\}$ satisfies $\sum_{j=0}^{\infty} a_j^2 < \infty$. We need this assumption to make sure that the variance of x_t is finite and, therefore, that x_t belongs to L^2 (proved later).

We call the stochastic process (1.5) an infinite order moving average and denote it by $MA(\infty)$. The Wold representation theorem (discussed below) proves that stochastic processes of the form (1.5) are sufficiently general to capture, in some sense to be discussed later, all linear properties of any covariance stationary stochastic process.

A special, but very important case of the process (1.5), is given by the family of ARMA models:

$$\begin{aligned} AR(p): &\quad x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \varepsilon_t \\ MA(q): &\quad x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} \\ ARMA(p,q): &\quad x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}. \end{aligned}$$

[1] $\gamma(\tau) = E[(X_t - \mu)(X_{t-\tau} - \mu)] = E[(X_{t+\tau} - \mu)(X_t - \mu)] = E\left[(X_t - \mu)\left(X_{t-(-\tau)} - \mu\right)\right] = \gamma(-\tau)$, where the second equality uses stationarity.

Lag operators

Once we start taking linear combinations of current and lagged variables, algebra becomes messy pretty fast. Therefore, it is useful to introduce the concept of the **lag operator**. The lag operator takes a sequence as input and delivers another sequence as output which is equal to the original sequence with the index lagged one period. That is $L\{x_t\} = \{y_t\}$ where $y_t = x_{t-1}$. But to avoid clutter, we simply write

$$Lx_t = x_{t-1}.$$

Clearly, $L(Lx_t) = Lx_{t-1} = x_{t-2}$. Let $L^2 x_t$ denote this double application of the lag operator. More generally, $L^p x_t = x_{t-p}$ for any $p \geq 1$. We also have $L^{-p} x_t = x_{t+p}$ and this defines the forward operator.[2]

We can also *define* a polynomial in the lag operator $a(L)$ as

$$a(L) = a_0 + a_1 L + a_2 L^2 + ... = \sum_{j=1}^{\infty} a_j L^j,$$

where $L^0 \equiv 1$. With this notation, the process (1.5) can be written as

$$y_t = a(L)\varepsilon_t = \left(\sum_{j=0}^{\infty} a_j L^j \right) \varepsilon_t.$$

ARMA models can be written in terms of the lag operator as

$$\begin{array}{rc} AR(p): & (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) x_t = \varepsilon_t \\ MA(q): & x_t = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q) \varepsilon_t \\ ARMA(p,q): & (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) x_t = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q) \varepsilon_t. \end{array}$$

We manipulate lag polynomials as if they were regular polynomials. For example, take an AR(1) model. By repeated substitution we have

$$\begin{aligned} x_t =& \phi x_{t-1} + \varepsilon_t \\ =& \phi^2 x_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ =& \phi^3 x_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ & \vdots \\ =& \phi^{s+1} x_{t-s-1} + \phi^s \varepsilon_{t-s} + \phi^{s-1} \varepsilon_{t-s+1} + ... + \phi \varepsilon_{t-1} + \varepsilon_t, \end{aligned}$$

so that, if $|\phi| < 1$, the term $\phi^{s+1} x_{t-s-1}$ tends to zero in the mean-squared sense:

$$\lim_{s \to \infty} E\left(\phi^{s+1} x_{t-s-1}\right)^2 = \lim_{s \to \infty} \phi^{2(s+1)} E\left(x^2\right) = E\left(x^2\right) \lim_{s \to \infty} \phi^{2(s+1)} = 0.$$

[2] Formally, if we define the forward operator as $L^{-1} x_t = x_{t+1}$, it is easy to see that L^{-1} is the inverse of the lag operator L—and hence the notation. Indeed, recall that a function $h(\cdot)$ is called the inverse of a function $f(\cdot)$ if $h(f(x)) = x$. Therefore, we immediately see that $L^{-1}(Lx_t) = L^{-1}(x_{t-1}) = x_t$. This proves that the forward operator is the inverse of the lag-operator.

Taking the limit as $s \to \infty$ we obtain the $MA(\infty)$ representation of the AR(1) model

$$x_t = \sum_{s=0}^{\infty} \phi_s \varepsilon_{t-s} = \left[1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + ...\right] \varepsilon_t,$$

where $\phi_0 = 1$. Note that we can obtain the same expression using the lag operator as follows: write the AR(1) model as

$$(1 - \phi L)\, x_t = \varepsilon_t.$$

We need to *invert* the lag polynomial $(1 - \phi L)$. That is, we want to find an operator, which we denote by $(1 - \phi L)^{-1}$ or $1/(1 - \phi L)$, such that $(1 - \phi L)^{-1}(1 - \phi L) = 1$.

Recall that, for a real number c with $|c| < 1$, we have the geometric series expansion

$$\frac{1}{1-c} = 1 + c + c^2 + c^3 + \ldots.$$

This expansion *suggests* treating ϕL like a real number, with the hope that $|\phi| < 1$ implies $|\phi L| < 1$ in some sense. If this interpretation is correct[3], we obtain

$$(1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \ldots.$$

Therefore,

$$x_t = \frac{\varepsilon_t}{1 - \phi L} = \left[1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + ...\right] \varepsilon_t.$$

Multiplication of lag polynomials works in the obvious way. Suppose $a(L) = a_0 + a_1 L + a_2 L^2$ and $b(L) = b_0 + b_1 L + b_2 L^2$, then

$$\begin{aligned} a(L)\, b(L) &= \left(a_0 + a_1 L + a_2 L^2\right)\left(b_0 + b_1 L + b_2 L^2\right) \\ &= a_0 b_0 + a_0 b_1 L + a_0 b_2 L^2 + b_0 a_1 L + b_1 a_1 L^2 + b_2 a_1 L^3 + b_0 a_2 L^2 + b_1 a_2 L^3 + b_2 a_2 L^4 \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0) L + (a_0 b_2 + b_1 a_1 + b_0 a_2) L^2 + (b_2 a_1 + b_1 a_2) L^3 + b_2 a_2 L^4 \end{aligned}$$

Here is another trick for lag operators. Let's write the following AR(2) model,

$$\left(1 - \phi_1 L - \phi_2 L^2\right) x_t = \varepsilon_t,$$

in terms of a $MA(\infty)$ representation. Rather than inverting the AR(2) by brute force, we write

$$\begin{aligned} 1 - \phi_1 L - \phi_2 L^2 &= (1 - \lambda_1 L)(1 - \lambda_2 L) \\ &= 1 - (\lambda_1 + \lambda_2) L + \lambda_1 \lambda_2 L^2. \end{aligned}$$

Thus, λ_1 and λ_2 solve

$$\lambda_1 + \lambda_2 = \phi_1;\ \lambda_1 \lambda_2 = -\phi_2.$$

Therefore,

$$(1 - \lambda_1 L)(1 - \lambda_2 L)\, x_t = \varepsilon_t.$$

[3] And it is indeed correct. But to make this statement formal we need to define and analyze properties of the $z-$transform. See, for example, Gabel and Roberts, chapter 4.

These polynomials are invertible as long as $|\lambda_1| < 1$ and $|\lambda_2| < 1$. If this is the case, we can write

$$\begin{aligned} x_t &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\varepsilon_t \\ &= \left(\sum_{j=0}^{\infty}\lambda_1^j L^j\right)\left(\sum_{i=0}^{\infty}\lambda_2^i L^i\right)\varepsilon_t. \end{aligned}$$

This is still ugly. If $\lambda_1 \neq \lambda_2$, we can use another trick: partial fractions expansions

$$\begin{aligned} \frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)} &= \frac{a}{1-\lambda_1 L}+\frac{b}{1-\lambda_2 L} \\ &= \frac{a(1-\lambda_2 L)+b(1-\lambda_1 L)}{(1-\lambda_1 L)(1-\lambda_2 L)} \\ &= \frac{a+b-(a\lambda_2+b\lambda_1)L}{(1-\lambda_1 L)(1-\lambda_2 L)}, \end{aligned}$$

which is true as long as $a+b=1$ and $a\lambda_2+b\lambda_1=0$ or

$$a=\frac{\lambda_1}{\lambda_1-\lambda_2},\ b=\frac{\lambda_2}{\lambda_2-\lambda_1}.$$

Therefore,

$$\begin{aligned} x_t &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\varepsilon_t \\ &= \left[\left(\frac{\lambda_1}{\lambda_1-\lambda_2}\right)\frac{1}{1-\lambda_1 L}+\left(\frac{\lambda_2}{\lambda_2-\lambda_1}\right)\frac{1}{1-\lambda_2 L}\right]\varepsilon_t \\ &= \sum_{j=0}^{\infty}\left[\frac{\lambda_1}{\lambda_1-\lambda_2}\lambda_1^j+\frac{\lambda_2}{\lambda_2-\lambda_1}\lambda_2^j\right]\varepsilon_{t-j}. \end{aligned}$$

When $\lambda_1=\lambda_2$, the algebra is different and we have to use binomial expansions with negative exponents (see Sargent (1987), pages 194-195).

In general, for an $AR(p)$ process we need to find the p roots of the polynomial $1-\phi_1 z-\phi_2 z^2-...-\phi_p z^p=0$. The λ's of the previous decomposition are the reciprocal of these roots. The $AR(p)$ is invertible as long as all roots of the above polynomial are greater than 1 in absolute value (so that the reciprocal of the roots, or the λ's, are *less* than one in absolute value). In this case we can write the $AR(p)$ model as

$$\begin{aligned} y_t &= \phi(L)^{-1}\varepsilon_t = \left[1-\phi_1 L-\phi_2 L^2-...-\phi_p L^p\right]^{-1}\varepsilon_t \\ &= \left[(1-\lambda_1 L)(1-\lambda_2 L)...(1-\lambda_p L)\right]^{-1}\varepsilon_t. \end{aligned}$$

If all the λ's are different, the partial fractions trick implies

$$\frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)...(1-\lambda_p L)}=\sum_{i=1}^{p}\frac{a_i}{1-\lambda_i L},$$

where

$$a_i=\frac{\lambda_i}{\prod_{j\neq i}(\lambda_i-\lambda_j)}\text{ for all } i,$$

so that

$$x_t = \sum_{j=0}^{\infty} \left(\sum_{i=1}^{p} a_i \lambda_i^j \right) \varepsilon_{t-j}.$$

More tricks using the lag operator can be found in Sargent (1987) and Cochrane (2005).

We now consider conditions under which a MA process is weakly stationary. We need to show that unconditional means and covariances are finite and do not depend on time. Write the $MA(\infty)$ as

$$x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$$

where ε_t satisfies $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2$ and $E[\varepsilon_t \varepsilon_{t-j}] = 0$ for $j \neq 0$. First, note that $E[x_t] = \sum_{j=0}^{\infty} \theta_j E[\varepsilon_t] = 0$, so the unconditional mean does not depend on time. The variance of x_t is given by

$$E\left[x_t^2\right] = E[\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}]^2 = \sum_{j=0}^{\infty} \theta_j^2 E\left[\varepsilon_{t-j}^2\right] = \sigma^2 \sum_{j=0}^{\infty} \theta_j^2$$

which is finite if and only if $\sum_{j=0}^{\infty} \theta_j^2 < \infty$.

The autocovariance $\gamma(\tau)$ is

$$\begin{aligned} E[x_t x_{t-\tau}] &= E\left[\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \sum_{h=0}^{\infty} \theta_h \varepsilon_{t-\tau-h}\right] = \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \theta_j \theta_h E[\varepsilon_{t-j} \varepsilon_{t-\tau-h}] \\ &= \sum_{j=0}^{\infty} \theta_j \theta_{j-\tau} E[\varepsilon_{t-j} \varepsilon_{t-j}] = \sigma^2 \sum_{j=0}^{\infty} \theta_j \theta_{j-\tau}. \end{aligned}$$

From the second to the third equality we use that $E[\varepsilon_{t-j} \varepsilon_{t-\tau-h}] = 0$ for all $j \neq \tau - h$ and $E[\varepsilon_{t-j} \varepsilon_{t-\tau-h}] = \sigma^2$ for $j = h + \tau \Rightarrow h = j - \tau$. This proves that $\gamma(\tau)$ depends only on τ and not on t. Moreover, by the Cauchy-Schwarz inequality we know that $|\gamma(\tau)| \leq \gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \theta_j^2$ which implies that all autocovariances are finite. Thus, the $MA(\infty)$ is covariance stationary as long as $\sum_{j=0}^{\infty} \theta_j^2 < \infty$.

Remark: One has to be careful when passing the expectation operator through the summation operator—and also, when we claimed above that in the AR(1) case, $\lim_{s \to \infty} \phi^s x_{t-s} = 0$. This requires using some theorems from the Lebesgue theory of integration which are beyond the scope of these notes. But if are a nerd, see pages 26-39 of Fuller (1995).

Linear projections

Here we consider a problem that will appear in various forms throughout the course. Let y, x_1, x_2, ... , x_n denote a set of random variables in L^2. The Cauchy-Schwarz inequality (1.3) implies that the second moments $E[yx_j]$ and $E[x_i x_j]$ exist and are finite as well. Moreover, it can be shown that, if second moments exist, first moments exist as well, so $E[y]$, $[x_1]$, $E[x_2]$, ..., $E[x_n]$ are also finite.

Consider estimating the random variable y on the basis of knowing the values of the random variables x_1, x_2,..., x_n. In particular, we want to compute the **best linear projection** defined as the linear (affine) function

$$\hat{y} = a_0 + a_1 x_1 + ... + a_n x_n$$

that best approximates y.

By best approximation we mean the following: we choose numbers a_i that makes the random variable $\hat{y}$ as close as possible to y in the least squares sense $E\left(y - \hat{y}\right)^2$:

$$\min_{\{a_i\}} E\left[\left(y - a_0 x_0 - a_1 x_1 - ... - a_n x_n\right)^2\right], \tag{1.6}$$

where we created a trivial random variable $x_0 \equiv 1$. This is equivalent to ordinary least squares but using population rather than sample moments.

Theorem 1.1 — Orthogonality principle

The numbers $a_0, a_1, a_2..., a_n$ minimize (1.6) if and only if

$$E\left[\left(y - a_0 x_0 - a_1 x_1 - ... - a_n x_n\right) x_i\right] = 0 \text{ for } i = 0, 1, 2, ..., n. \tag{1.7}$$

Proof. Let $a = (a_0, a_1, ..., a_n)'$ and consider the minimization problem

$$\min_a J(a) = \min_a \frac{1}{2} E[(y - \sum_{j=0}^{n} a_j x_j)^2].$$

Differentiating (1.6) with respect to a_i gives

$$\frac{\partial J(a)}{\partial a_i} = -E(y - \sum_{j=0}^{n} a_j x_j) x_i = 0 \text{ for } i = 0, 1, 2, ..., n.$$

This shows that (1.7) is a necessary condition. If we show that the problem is strictly convex, the minimizer is unique and sufficiency also holds.

Let $x = (x_0, x_1, x_2, ..., x_n)'$ be an $(n+1) \times 1$ column vector. The first order conditions can be written in matricial form as

$$\nabla_a J(a) = -\left[E(xy) - E(xx')\, a\right] = \mathbf{0}_{n+1 \times 1}.$$

Differentiating this expression with respect to a' gives

$$\nabla_{aa'} J(a) = E(xx'),$$

which is positive definite because $E[xx']$ is a covariance matrix. Therefore, the minimization problem is convex and the first order conditions are sufficient. This completes the proof. ■

Assuming that $E[xx']$ is invertible[4], we can obtain the optimal weights a by solving

$$a = E(xx')^{-1} E(xy). \tag{1.8}$$

The random variable $\sum_{j=0}^{n} a_j x_j$ is called the **projection** of y onto $\{1, x_1, x_2, ..., x_n\}$.

The orthogonality principle implies that the projection error, $y - \sum_{j=0}^{n} a_j x_j$, is **orthogonal** to each of the x_i and, therefore, to any linear combination of them. (Two random variables x, y are orthogonal if $E[xy] = 0$.) Defining the projection error as ε, it follows that

$$y = \sum_{j=0}^{n} a_j x_j + \varepsilon \tag{1.9}$$

where $E[\varepsilon \sum_{i=0}^{n} \phi_i x_i] = 0$ for any $\{\phi_i\}$ (*why?*). Thus, equation (1.9) decomposes y into two orthogonal components: $\sum_{j=0}^{n} a_j x_j$ and ε.

It then follows that

$$E\left(y^2\right) = E\left(\sum_{j=0}^{n} a_j x_j\right)^2 + E\left(\varepsilon^2\right).$$

In addition, note that $E(\varepsilon) = 0$. This follows from the orthogonality condition for $i = 0$ above. The key for this result is to include a constant in the projection; without the constant, the forecast error ε need not have zero mean.

Sometimes we use the following notation for the projection

$$P[y|1, x_1, x_2, ..., x_n] \equiv x'a = \sum_{j=0}^{n} a_j x_j.$$

Lemma. *The projection is a linear operator:*

$$P[\alpha y + \beta z|1, x_1, x_2, ..., x_n] = \alpha P[y|1, x_1, x_2, ..., x_n] + \beta P[z|1, x_1, x_2, ..., x_n].$$

Proof. Let $P[y|1, x_1, x_2, ..., x_n] = \sum_{j=0}^{n} a_j x_j$ and $P[z|1, x_1, x_2, ..., x_n] = \sum_{j=0}^{n} b_j x_j$. The orthogonality principle implies

$$\begin{aligned} E\left(y - \sum_{j=0}^{n} a_j x_j\right) x_i &= 0 \text{ for all } i \\ E\left(z - \sum_{j=0}^{n} b_j x_j\right) x_i &= 0 \text{ for all } i \end{aligned}$$

[4] $E[xx']$ not invertible means that (at least) one of the random variables x_i is a linear combination of the others. If a random variables is a linear combinations of the others it does not add anything to the linear projection. We can simply delete these variables until we obtain an invertible covariance matrix $E[xx']$.

Multiplying the first condition by α and the second by β gives

$$\begin{aligned} E\left(\alpha y - \alpha \sum_{j=0}^{n} a_j x_j\right) x_i &= 0 \text{ for all } i \\ E\left(\beta z - \beta \sum_{j=0}^{n} b_j x_j\right) x_i &= 0 \text{ for all } i \end{aligned}$$

Adding these equations gives

$$E\left[\alpha y + \beta z - \sum_{j=0}^{n} (\alpha a_j + \beta b_j) x_j\right] x_i = 0 \text{ for all } i.$$

This means that the numbers $(\alpha a_j + \beta b_j)$ for $j = 0, 1, 2, ..., n$ satisfy the orthogonality principle of a projection of $\alpha y + \beta z$ onto $\{1, x_1, x_2, ..., x_n\}$. Therefore,

$$\begin{aligned} P[\alpha y + \beta z | 1, x_1, x_2, ..., x_n] &= \sum_{j=0}^{n} (\alpha a_j + \beta b_j) x_j \\ &= \alpha \sum_{j=0}^{n} a_j x_j + \beta \sum_{j=0}^{n} b_j x_j \\ &= \alpha P[y | 1, x_1, x_2, ..., x_n] + \beta P[z | 1, x_1, x_2, ..., x_n], \end{aligned}$$

which completes the proof. ■

Recursive projections

Here we show how to update a projection when new information arrives. This result will be useful to derive the updating formula for the Kalman filter.

We observe a set of random variables $\Omega = \{1, x_1, x_2, ..., x_n\}$ (we include the constant 1 in Ω) and compute the projection $P[y|\Omega]$. Suppose that we are given a new set of random variables $\mathbf{z} = (z_1, z_2, ..., z_m)'$ and want to compute (update) the linear projection $P[y|\Omega, \mathbf{z}]$ based on our knowledge of $P[y|\Omega]$.

Consider the decomposition (1.9) for the updated projection:

$$\begin{aligned} y =& P[y|\Omega, \mathbf{z}] + \varepsilon \\ =& \sum_{j=0}^{n} a_j x_j + \sum_{s=1}^{m} \delta_s z_s + \varepsilon. \end{aligned} \tag{1.10}$$

where $E(\varepsilon) = 0$, $E(\varepsilon x_j) = 0$ for $j = 1, 2, ..., n$, and $E(\varepsilon z_s) = 0$ for $s = 1, 2, ...m$. The orthogonality principle guarantees that the a_j's and δ_s's are the least square parameter values. Now project both sides of (1.10) on the smaller set Ω,

$$\begin{aligned} P[y|\Omega] =& P\left[\sum_{j=0}^{n} a_j x_j + \sum_{s=1}^{m} \delta_s z_s + \varepsilon | \Omega\right] \\ =& \sum_{j=0}^{n} a_j P[x_j|\Omega] + \sum_{s=1}^{m} \delta_s P[z_s|\Omega] + P[\varepsilon|\Omega]. \end{aligned}$$

where we used that the projection is a linear operator. Moreover, $P[x_j|\Omega] = x_j$ for $j = 1, 2, ..., n$ and $P[\varepsilon|\Omega] = 0$. To see the former, consider the objective function of the least square problem for

$$\min_{a_0,a_1,..,a_n} E\left[(x_j - a_0x_0 - a_1x_1 - a_2x_2 - ... - a_nx_n)^2\right].$$

This is clearly minimized when $a_j = 1$ and $a_i = 0$ for all $i \neq j$. To see the latter, note that the orthogonality conditions of (1.10) imply that all the coefficients of the linear projection of ε on x_j for $j = 1, 2, ..., n$ are zero. Thus, we obtain

$$P[y|\Omega] = \sum_{j=0}^{n} a_j x_j + \sum_{s=1}^{m} \delta_s P[z_s|\Omega]. \tag{1.11}$$

Subtracting (1.11) from (1.10) then gives

$$y - P[y|\Omega] = \sum_{s=1}^{m} \delta_s (z_s - P[z_s|\Omega]) + \varepsilon. \tag{1.12}$$

This equation looks like a projection of $y - P[y|\Omega]$ on $z_s - P[z_s|\Omega]$. To confirm this conjecture, we need to show that ε is orthogonal to $z_s - P[z_s|\Omega]$ for all s. But this is obvious because ε is orthogonal to z_s and x_j, and $P[z_s|\Omega]$ is a linear function of x_j, hence $E[(z_s - P[z_s, \Omega])\varepsilon] = 0$ for all s. Therefore, the orthogonality principle implies that δ_s for $s = 1, 2, ..., m$ are the coefficients of the projection of $(y - P[y|\Omega])$ on $(z_s - P[z_s, \Omega])$,

$$P[(y - P[y|\Omega]) \,|\, (\mathbf{z} - P[\mathbf{z}|\Omega])] = \sum_{s=1}^{m} \delta_s (z_s - P[z_s|\Omega]).$$

Rearranging (1.12) and using the previous result gives

$$y = P[y|\Omega] + P[(y - P[y|\Omega]) \,|\, (\mathbf{z} - P[\mathbf{z}|\Omega])] + \varepsilon.$$

Because ε is orthogonal to $\{\Omega, \mathbf{z}\}$, it then follows that

$$P[y|\Omega, z] = \underbrace{P[y|\Omega]}_{\text{Original projection}} + \underbrace{P[(y - P[y|\Omega]) \,|\, (\mathbf{z} - P[\mathbf{z}|\Omega])]}_{\text{Projection of prediction errors on prediction errors}}. \tag{1.13}$$

In words, to update a linear projection when new information $\mathbf{z}$ arrives, one adds to the original projection, $P[y|\Omega]$, the projection of the prediction errors of the original projection, $y - P[y|\Omega]$, on the prediction errors of the projection of the new variables $\mathbf{z}$ on the original set of variables, $\mathbf{z} - P(\mathbf{z}|\Omega)$.

For future use, let's write (1.13) in vector notation. Let $x = (1, x_1, x_2, ..., x_n)'$ and $z = (z_1, z_2, ..., z_m)'$. Furthermore, stack the projection coefficients as $a = (a_0, a_1, ..., a_n)'$ and $b^s = (b_0^s, b_1^s, b_2^s, ..., b_n^s)'$ for all s. The normal equations are

$$\begin{aligned} a =& E(xx')^{-1} E(xy) \\ b^s =& E(xx')^{-1} E(xz_s) \end{aligned}$$

which implies

$$\begin{aligned} P[y|x] =& x'a = x'E(xx')^{-1}E(xy) \\ P[z_s|x] =& x'b^s = x'E(xx')^{-1}E(xz_s). \end{aligned}$$

The projection errors are thus

$$y - P[y|x] = y - x'a = y - x'E(xx')^{-1}E(xy)$$

$$u_s \equiv z_s - P[z_s|x] = z_s - x'b^s = z_s - x'E(xx')^{-1}E(xz_s).$$

Let $u = (u_1, u_2, ..., u_m)'$ denote the vector of the projections errors of z_s on x for $s = 1, 2, ..., m$, and $B = [b^1\ b^2\ ...\ b^m]$ be the $(n+1) \times m$ matrix whose column s contains the vector of projection coefficients of z_s on x. Then, we can write

$$u = z - B'x.$$

It then follows that the $m \times 1$ vector δ of coefficients of the projection of the forecast errors $y - P[y|x]$ on the forecast errors u is given by

$$\begin{aligned} \delta =& E(uu')^{-1}E[u(y - P[y|x])] \\ =& E(uu')^{-1}E\left[u\left(y - x'E[xx']^{-1}E[xy]\right)\right] \end{aligned}$$

so that

$$P\left[\left(y - x'E[xx']^{-1}E[xy]\right)|u\right] = u'\delta = u'E(uu')^{-1}E\left[u\left(y - x'E(xx')^{-1}E(xy)\right)\right].$$

Therefore

$$\begin{aligned} P[y|x,z] =& P[y|x] + P[(y - P[y|x])\,|\,(z - P[z|x])] \\ =& x'E(xx')^{-1}E(xy) + u'E(uu')^{-1}E\left[u\left(y - x'E(xx')^{-1}E(xy)\right)\right], \end{aligned}$$

where $u = z - B'x$.

Wold Representation Theorem

Above we constructed a covariance stationary process by combining white noise shocks

$$x_t = a(L)\varepsilon_t,$$

where $a(L) = 1 + a_1L + a_2L^2 + ...$ is a polynomial on the lag operator and $\sum_{j=0}^{\infty} a_j^2 < \infty$. The Wold representation theorem reverses the procedure: any weakly stationary process can be written as an infinite order moving average plus a perfectly predictable term.

Theorem 1.2 — Wold Representation Theorem

Any mean zero, covariance stationary process $\{x_t\}$ can be represented in the form

$$x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \eta_t \tag{1.14}$$

where

a) $\varepsilon_t = x_t - P[x_t|x_{t-1}, x_{t-2}, ...]$ is the prediction error of the projection of x_t on all its lags,

b) $P[\varepsilon_t|x_{t-1}, x_{t-2}, ...] = 0$, $E(\varepsilon_t x_{t-j}) = 0$ for all $j \geq 1$; $E(\varepsilon_t^2) = \sigma^2$ for all t; $E(\varepsilon_t) = 0$ for all t; $E(\varepsilon_t \varepsilon_s) = 0$ for all $s \neq t$,

c) $\theta_0 = 1$; $\sum_{j=0}^{\infty} \theta_j^2 < \infty$,

d) $\{\theta_j\}$ and $\{\varepsilon_t\}$ are unique,

e) η_t is linearly deterministic: $\eta_t = P[\eta_t|x_{t-1}, x_{t-2}, x_{t-3}, ...]$.

Before going to the proof of the theorem, we mention what the theorem says and what the theorem does not say (mostly from Cochrane, 2005):

a) The ε_t's are a white noise but need not be i.i.d. or normally distributed.

b) Although $E(\varepsilon_t x_{t-j}) = 0$ (ε_t and x_{t-j} are orthogonal) $E(\varepsilon_t|x_{t-j})$ need not be zero. This is the difference between orthogonality and independence: two random variables x and y can be orthogonal but not independent. For example, let x be normal with mean zero and variance σ^2 and let $y = x^2$. Then $E(xy) = E(x^3) = 0$ but $E(y|x) = x^2$.

c) The innovations ε_t are prediction errors. They do not have a structural interpretation as the shocks of a model. The Wold decomposition is a probabilistic decomposition.

d) The Wold decomposition is *one linear representation* of the process $\{x_t\}$. There could be other non-linear representations that may be better in some sense. Moreover, the Wold decomposition is not even the *unique linear* $MA(\infty)$ representation of the process (see below).

e) We usually ignore η_t.

We provide a sketch of the proof following Sargent (1987)–the formal proof requires being more careful in some steps.

Proof of the Wold Representation Theorem. The proof is constructive. We divide the proof in a number of steps:

Step 1: construct the white noise process ε_t.

Using the orthogonality principle, write

$$x_t = P\left[x_t | x_{t-1}, x_{t-2}, ...\right] + \varepsilon_t,$$

where $P\left[x_t | x_{t-1}, x_{t-2}, ...\right]$ is the projection of x_t on the entire history of past x's, and ε_t is a prediction error orthogonal to x_{t-j} for $j = 1, 2, ...$ This defines the sequence of unique innovations $\{\varepsilon_t\}$ because the projection is unique. Furthermore, since each x_t has mean zero and ε_t is a linear combination of x's, then $E(\varepsilon_t) = 0$.

Let σ^2 be the mean squared error of the projection,

$$\sigma^2 = E\left(\varepsilon_t^2\right) = E\left(x_t - P\left[x_t | x_{t-1}, x_{t-2}, ...\right]\right)^2.$$

Note that σ^2 does not depend on t because ε_t is a linear combination of current and past x's, and x is covariance stationary.

Using that

$$\varepsilon_{t-s} = x_{t-s} - P\left[x_{t-s} | x_{t-s-1}, x_{t-s-2}, ...\right]$$

is a linear combination of $x_{t-s}, x_{t-s-1}, ...$ and that the orthogonality principle implies $E\left[\varepsilon_t x_{t-s}\right] = 0$ for all $s \geq 1$, it follows that $E\left[\varepsilon_t \varepsilon_{t-s}\right] = 0$ for all $s \geq 1$. This proves that $\{\varepsilon_t\}$ is a serially uncorrelated process.

Step 2: Construct the coefficients θ_j of the projection of x_t on past ε

Consider projecting x_t on a sequence of (finite) sets $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-m}\}$ for successively larger m's. Denote the projection of x_t on such set as

$$\hat{x}_t^m = P\left[x_t | \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-m}\right] = \sum_{j=0}^{m} \theta_j \varepsilon_{t-j}.$$

The orthogonality principle implies that the prediction error is orthogonal to each ε in the set,

$$E[(x_t - \sum_{j=0}^{m} \theta_j \varepsilon_{t-j})\varepsilon_{t-k}] = 0 \text{ for } k = 0, 1, 2, ..., m.$$

Since $E\left[\varepsilon_{t-j}\varepsilon_{t-k}\right] = 0$ for all $j \neq k$ we have

$$E\left[x_t \varepsilon_{t-k}\right] - \theta_k E\left[\varepsilon_{t-k}^2\right] = 0 \text{ for } k = 0, 1, 2, ..., m$$

so that

$$\theta_k = \frac{E\left(x_t \varepsilon_{t-k}\right)}{\sigma^2} \text{ for } k = 0, 1, 2, ..., m.$$

Let $k = 0$. Since $\varepsilon_t = x_t - P\left[x_t | x_{t-1}, x_{t-2}, ...\right]$, it follows that $E\left(\varepsilon_t x_{t-j}\right) = 0$ for all $j \geq 1$ (orthogonality principle). Therefore,

$$\begin{aligned} E\left[\varepsilon_t x_t\right] &= E\left[\varepsilon_t \left(\varepsilon_t + P\left[x_t | x_{t-1}, x_{t-2}, ...\right]\right)\right] \\ &= E\left(\varepsilon_t^2\right) + \underbrace{E\left(\varepsilon_t P\left[x_t | x_{t-1}, x_{t-2}, ...\right]\right)}_{=0}. \end{aligned}$$

Thus, $\theta_0 = E\left(\varepsilon_t x_t\right) / E\left(\varepsilon_t^2\right) = E\left(\varepsilon_t^2\right) / E\left(\varepsilon_t^2\right) = 1$. *A key property of these projections is that* θ_k *does not depend on* m*, the length of the projection set* $\{\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_{t-m}\}$. This property reflects the lack of serial correlation of the ε's.

We now compute the variance of the prediction error

$$\begin{aligned}
0 \leq E\left(x_t - \sum_{j=0}^{m} \theta_j \varepsilon_{t-j}\right)^2 &= E[x_t^2 - 2x_t \sum_{j=0}^{m} \theta_j \varepsilon_{t-j} + (\sum_{j=0}^{m} \theta_j \varepsilon_{t-j})^2] \\
&= E\left(x_t^2\right) - 2\sum_{j=0}^{m} \theta_j E\left(x_t \varepsilon_{t-j}\right) + \sum_{j=0}^{m} \theta_j^2 E\left(\varepsilon_{t-j}\right)^2 \\
&= E\left(x_t^2\right) - 2\sum_{j=0}^{m} \theta_j \sigma^2 \frac{E\left(x_t \varepsilon_{t-j}\right)}{\sigma^2} + \sum_{j=0}^{m} \theta_j^2 \sigma^2 \\
&= E\left(x_t^2\right) - 2\sigma^2 \sum_{j=0}^{m} \theta_j^2 + \sigma^2 \sum_{j=0}^{m} \theta_j^2 \\
&= E\left(x_t^2\right) - \sigma^2 \sum_{j=0}^{m} \theta_j^2,
\end{aligned}$$

where the second equality follows because the ε_{t-j} are uncorrelated, and the fourth equality uses the definition of θ_j. Since $E\left(x_t^2\right) < \infty$, the previous inequality implies

$$\sum_{j=0}^{m} \theta_j^2 \leq \frac{E\left(x_t^2\right)}{\sigma^2} < \infty \text{ for all } m.$$

Taking the limit as $m \to \infty$ proves that $\sum_{j=0}^{\infty} \theta_j^2 < \infty$—the sequence $\{\theta_j\}$ is square summable.

The square summability of $\{\theta_j\}$ implies that the projection $\hat{x}_t^m$ is a Cauchy sequence. To see this, take $n > m$ and compute

$$\begin{aligned}
\|\hat{x}_t^n - \hat{x}_t^m\|^2 &= E\left(\hat{x}_t^n - \hat{x}_t^m\right)^2 \\
&= E\left(\sum_{j=0}^{n} \theta_j \varepsilon_{t-j} - \sum_{j=0}^{m} \theta_j \varepsilon_{t-j}\right)^2 \\
&= E\left(\sum_{j=m+1}^{n} \theta_j \varepsilon_{t-j}\right)^2 \\
&= \sum_{j=m+1}^{n} \theta_j^2 \sigma^2 \leq \sigma^2 \sum_{j=m+1}^{\infty} \theta_j^2
\end{aligned}$$

Since $\sum_{j=0}^{m} \theta_j^2 < \infty$ for all m, it follows that we can take m big enough to make $\sum_{j=m+1}^{\infty} \theta_j^2$ arbitrarily close to zero. This means that $\hat{x}_t^m$ is Cauchy. Therefore, the completeness of the Hilbert space L^2 implies that there exist an element $\hat{x}_t \in L_2$ such that

$$\hat{x}_t^m \to \hat{x}_t \equiv \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}. \tag{1.15}$$

Step 3: Construct the component η_t

Let η_t be the difference between x_t and the projection of x_t onto the current and past ε_t's

$$\eta_t \equiv x_t - \hat{x}_t = x_t - \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \tag{1.16}$$

We first establish that $E(\eta_t \varepsilon_s) = 0$ for all s and t. It should be clear that $E(\eta_t \varepsilon_s) = 0$ for $s > t$ because η_t is a linear function of x_t and the sequence of shocks $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...$ However, ε_s is orthogonal to all other ε_t's and to x_t when $s > t$. Consider now the case $s \leq t$,let $s = t - k$ for the appropriate $k \geq 0$ and compute

$$\begin{aligned} E(\eta_t \varepsilon_{t-k}) &= E(x_t \varepsilon_{t-k}) - E(\hat{x}_t \varepsilon_{t-k}) \\ &= E(x_t \varepsilon_{t-k}) - E[(\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}) \varepsilon_{t-k}] \\ &= \sigma^2 \theta_k - \sigma^2 \theta_k = 0. \end{aligned}$$

where we are using the definition of θ_k. Thus, the entire $\{\varepsilon_t\}$ process is orthogonal to the process $\{\eta_t\}$.

We next establish that η_t is perfectly predictable from past observations of x_t's. In particular, project η_t on $\{x_{t-1}, x_{t-2}, ...\}$

$$P[\eta_t | x_{t-1}, x_{t-2}, ...] = P[x_t | x_{t-1}, x_{t-2}, ...] - \sum_{j=0}^{\infty} \theta_j P[\varepsilon_{t-j} | x_{t-1}, x_{t-2}, ...], \tag{1.17}$$

where we used linearity of the projection. Consider the second term on the right hand side. In the case $j = 0$ we need to compute $P[\varepsilon_t | x_{t-1}, x_{t-2}, ...]$. However, we already established that ε_t is orthogonal to all past x_t's, so that $E(\varepsilon_t x_{t-j}) = 0$ for all $j \geq 1$ which implies

$$P[\varepsilon_t | x_{t-1}, x_{t-2}, ...] = 0.$$

Consider now computing the projection $P[\varepsilon_t | x_t, x_{t-1}, x_{t-2}....]$. By step 1 above, ε_t is a linear combination of the current and past x_t's. Therefore, the orthogonality principle implies $P[\varepsilon_t | x_t, x_{t-1}....] = \varepsilon_t$ because, ε_t being a linear combination of current and past x_t's, means that we can set the objective function of the projection problem to exactly zero. A similar argument can be made to argue that

$$P[\varepsilon_t | x_{t+j-1}, x_{t+j-2}, ...] = \varepsilon_t \text{ for } j \geq 0.$$

In effect, ε_t is a linear combination of a subset of the projecting variables $x_{t+j-1}, x_{t+j-2}, ...$ It will then be possible to make the objective function of the projection exactly equal to zero as well.

This implies that

$$\begin{aligned} P[\varepsilon_t | x_{t-1}, x_{t-2}, ...] &= \varepsilon_t \\ P[\varepsilon_{t-1} | x_{t-1}, x_{t-2}, ...] &= \varepsilon_{t-1} \\ &\vdots \\ P[\varepsilon_{t-j} | x_{t-1}, x_{t-2}, ...] &= \varepsilon_{t-j}. \end{aligned}$$

Therefore, (1.17) becomes

$$P[\eta_t | x_{t-1}, x_{t-2}, ...] = P[x_t | x_{t-1}, x_{t-2}, ...] - \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}.$$

Subtracting (1.16) from this expression gives

$$\begin{aligned}
\eta_t - P\left[\eta_t | x_{t-1}, x_{t-2}, ...\right] &= \left(x_t - \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}\right) - \left(P\left[x_t | x_{t-1}, x_{t-2}, ...\right] - \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}\right) \\
&= \underbrace{\left(x_t - P\left[x_t | x_{t-1}, x_{t-2}, ...\right]\right)}_{\varepsilon_t} - \left(\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} - \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}\right) \\
&= \underbrace{\left(x_t - P\left[x_t | x_{t-1}, x_{t-2}, ...\right]\right)}_{\varepsilon_t} - \theta_0 \varepsilon_t \\
&= \varepsilon_t - \theta_0 \varepsilon_t = 0.
\end{aligned}$$

All this algebra proves that

$$\eta_t = P\left[\eta_t | x_{t-1}, x_{t-2}, ...\right],$$

which means that the term η_t is "linearly deterministic" in the sense that it can be predicted without error using past information of x's. The remaining term, $\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$ is called the "linearly indeterministic" component of the process. This completes the proof of the theorem. ■

Wold representation for vector time series

The same decomposition holds for vector processes. Let $X_t = [x_{1t}, x_{2t}, ..., x_{nt}]'$ where each x_{it} is an individual stochastic process. We say that X_t is covariance stationary if $E[X_t] = \mu$ is independent of time and the matrix of autocovariances $E\left[(X_t - \mu)(X_{t-\tau} - \mu)'\right] = \Gamma_\tau$ only depends on τ and not on t.[5] As above, assume $\mu = 0$. Then, any covariance stationary vector process X_t can be represented as

$$X_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j} + \eta_t \tag{1.18}$$

where

a) $\varepsilon_t = X_t - P\left[X_t | X_{t-1}, X_{t-2}, X_{t-3}, ...\right]$ is the forecast error of the projection of the vector X_t on its lagged values,

b) $P\left[\varepsilon_t | X_{t-1}, X_{t-2}, X_{t-3}, ...\right] = 0$, $E(\varepsilon_t X_{t-j}) = 0$ for all $j \geq 1$; $E(\varepsilon_t^2) = \Sigma$ for all t is a constant covariance matrix; $E(\varepsilon_t) = 0$ for all t; $E(\varepsilon_t \varepsilon_s') = 0$ for all $s \neq t$,

c) Θ_j are $n \times n$ matrices that satisfy $\Theta_0 = I$; $\sum_{j=0}^{\infty} \Theta_j \Theta_j' < \infty$,

d) $\{\Theta_j\}$ and $\{\varepsilon_t\}$ are unique,

e) η_t is linearly deterministic; that is, $\eta_t = P\left[\eta_t | X_{t-1}, X_{t-2}, X_{t-3}, ...\right]$.

[5] Please note that the symmetry property now reads $\Gamma_\tau = \Gamma_{-\tau}'$.

An important remark: The Wold representation theorem shows that there is a unique representation of a covariance stationary process as a $MA(\infty)$ satisfying 1-5 above. This *does not mean* that (1.18) is the *unique* moving average representation of the process $\{X_t\}$. To see this, note that we can always write (1.18) as

$$X_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j} + \eta_t = \sum_{j=0}^{\infty} \Theta_j \Lambda \Lambda^{-1} \varepsilon_{t-j} + \eta_t = \sum_{j=0}^{\infty} \Phi_j \nu_{t-j} + \eta_t,$$

where Λ is an arbitrary $n \times n$ invertible matrix, $\Phi_j = \Theta_j \Lambda$ and $\nu_{t-j} = \Lambda^{-1}\varepsilon_{t-j}$. The innovation ν_t now satisfies $E(v_t) = \Lambda^{-1}E(\varepsilon_t) = 0$, $E\left(v_t v'_{t-s}\right) = 0$ for $s \neq 0$ and $E(v_t v'_t) = E\left(\Lambda^{-1}\varepsilon_t \varepsilon'_t (\Lambda^{-1})'\right) = \Lambda^{-1}\Sigma(\Lambda^{-1})'$. Therefore $X_t = \sum_{j=0}^{\infty} \Phi_j \nu_{t-j} + \eta_t$ is *another* infinite moving average representation of the vector process X_t. How do we relate this to the uniqueness claim in the Wold theorem? What happens here is that the residual ν_t *is not* the forecast error of projecting X_t on its infinite history.

This non-uniqueness result of the moving average representation of X_t will be used when discussing structural vector autoregressions later in the course.

Limit theorems

We use different versions of two limit theorems: Laws of Large Numbers (LLN) and Central Limit Theorems (CLT). Both are concerned with the behavior of sample means under different assumptions. The LLN is about convergence—in probability, almost surely, in L^2—of the sample mean to the population mean. The CLT is about convergence in distribution (the asymptotic distribution) of the sample mean. By appropriately weighting the sample mean by a function of the sample size (typically $\sqrt{T}$), the central limit theorem provides a non-degenerate distribution theory that can be used to test hypotheses, compute asymptotic confidence bands, etc.

Properties of the sample mean of a vector process (Hamilton, p. 279): Suppose that we have a sample of size T, $\{X_1, X_2, ..., X_T\}$ of an n dimensional vector process $\{X_t\}$, where X_t is covariance stationary with

$$\begin{aligned} E[X_t] &= \mu, \\ E\left[(X_t - \mu)(X_{t-v} - \mu)'\right] &= \Gamma_v. \end{aligned}$$

Assume also that the autocovariances are absolutely summable, that is $\sum_{v=-\infty}^{\infty} |\Gamma_v| < \infty$. If we let $\gamma_{ij}^{(v)}$ denote the element (i,j) of Γ_v, the requirement is that $\sum_{v=\infty}^{\infty} |\gamma_{ij}^{(v)}| = c_{ij} < \infty$. Recall also that for a vector process $\Gamma_v = \Gamma'_{-v}$.

Consider the sample mean

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t.$$

Clearly, $E\left[\bar{X}_T\right] = \frac{1}{T}\sum_{t=1}^{T} E[X_t] = \mu$.

The covariance matrix of the sample mean is

$$E\left[\left(\bar{X}_T-\mu\right)\left(\bar{X}_T-\mu\right)'\right]=E\left[\left(\frac{1}{T}\sum_{t=1}^{T}X_t-\mu\right)\left(\frac{1}{T}\sum_{t=1}^{T}X_t-\mu\right)'\right]$$

$$=\frac{1}{T^2}E\begin{bmatrix}(X_1-\mu)\left[(X_1-\mu)'+(X_2-\mu)'+(X_3-\mu)'+...+(X_T-\mu)'\right]+\\(X_2-\mu)\left[(X_1-\mu)'+(X_2-\mu)'+(X_3-\mu)'+...+(X_T-\mu)'\right]+\\(X_3-\mu)\left[(X_1-\mu)'+(X_2-\mu)'+(X_3-\mu)'+...+(X_T-\mu)'\right]+\\...+\\(X_T-\mu)\left[(X_1-\mu)'+(X_2-\mu)'+(X_3-\mu)'+...+(X_T-\mu)'\right]\end{bmatrix},$$

or

$$\begin{aligned}T^2E\left[\left(\bar{X}_T-\mu\right)\left(\bar{X}_T-\mu\right)'\right]=&\Gamma_0+\Gamma_{-1}+\Gamma_{-2}+...+\Gamma_{-(T-1)}+\\&\Gamma_1+\Gamma_0+\Gamma_{-1}+...+\Gamma_{-(T-2)}+\\&\Gamma_2+\Gamma_1+\Gamma_0+\Gamma_{-1}+...+\Gamma_{-(T-3)}+\\&+...+\\&\Gamma_{T-1}+\Gamma_{T-2}+...+\Gamma_0\\=&T\Gamma_0+(T-1)\Gamma_1+(T-1)\Gamma_{-1}+(T-2)\Gamma_2+(T-2)\Gamma_{-2}+\\&(T-3)\Gamma_3+(T-3)\Gamma_{-3}+...+(T-(T-1))\Gamma_{T-1}+(T-(T-1))\Gamma_{-(T-1)}\\=&\sum_{v=-(T-1)}^{T-1}(T-|v|)\Gamma_v.\end{aligned}$$

Thus

$$TE\left[\left(\bar{X}_T-\mu\right)\left(\bar{X}_T-\mu\right)'\right]=\sum_{v=-(T-1)}^{T-1}\left(1-\frac{|v|}{T}\right)\Gamma_v=\sum_{v=-(T-1)}^{T-1}\Gamma_v-\sum_{v=-(T-1)}^{T-1}\frac{|v|}{T}\Gamma_v.$$

The following is an important result that is used to prove the asymptotic distribution of several estimator,

Proposition 1:

$$\lim_{T\to\infty}TE\left[\left(\bar{X}_T-\mu\right)\left(\bar{X}_T-\mu\right)'\right]=\sum_{v=-\infty}^{\infty}\Gamma_v$$

Proof: Consider

$$\begin{aligned}\sum_{v=-\infty}^{\infty}\Gamma_v-TE\left[\left(\bar{X}_T-\mu\right)\left(\bar{X}_T-\mu\right)'\right]&=\sum_{v=-\infty}^{\infty}\Gamma_v-\sum_{v=-(T-1)}^{T-1}\Gamma_v+\sum_{v=-(T-1)}^{T-1}\frac{|v|}{T}\Gamma_v\\&=\sum_{|v|\geq T}\Gamma_v+\sum_{v=-(T-1)}^{T-1}\frac{|v|}{T}\Gamma_v.\end{aligned}$$

The (i,j) element of the above expression can be written as

$$\sum_{|v|\geq T}\gamma_{ij}^{(v)}+\sum_{v=-(T-1)}^{T-1}\frac{|v|}{T}\gamma_{ij}^{(v)}.$$

We need to prove that the absolute value of this term converges to zero for each i, j. Consider

$$\left| \sum_{|v|\geq T} \gamma_{ij}^{(v)} + \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T} \gamma_{ij}^{(v)} \right| \leq \sum_{|v|\geq T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| \tag{1.19}$$

Absolute summability of $\{\Gamma_v\}$ means that for any $\varepsilon > 0$ there exist an index q such that

$$\sum_{|v|>q}^{\infty} \left| \gamma_{ij}^{(v)} \right| < \frac{\varepsilon}{2},$$

for otherwise the sum will not converge. Now choose $T - 1 > q$ and write

$$\begin{aligned} \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| &= \sum_{v=-q}^{q} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=q+1}^{T-1} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=-(T-1)}^{-(q+1)} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| \\ &\leq \sum_{v=-q}^{q} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=q+1}^{T-1} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=-(T-1)}^{-(q+1)} \left| \gamma_{ij}^{(v)} \right| \end{aligned}$$

The inequality above uses that $|v|/T < 1$ for all $|v| = \{q+1, q+2, ..., T-1\}$.

Therefore, the right hand side of (1.19) can be written as

$$\begin{aligned} \sum_{|v|\geq T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| &\leq \sum_{|v|\geq T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=-q}^{q} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=q+1}^{T-1} \left| \gamma_{ij}^{(v)} \right| + \sum_{v=-(T-1)}^{-(q+1)} \left| \gamma_{ij}^{(v)} \right| \\ &= \sum_{v=-q}^{q} \frac{|v|}{T} \left| \gamma_{ij}^{(v)} \right| + \sum_{|v|>q} \left| \gamma_{ij}^{(v)} \right| \\ &< \frac{1}{T} \sum_{v=-q}^{q} |v| \left| \gamma_{ij}^{(v)} \right| + \frac{\varepsilon}{2}. \end{aligned}$$

where the last inequality uses that q satisfies the inequality $\sum_{|v|>q} \left| \gamma_{ij}^{(v)} \right| < \varepsilon/2$. Putting together these inequalities gives

$$\left| \sum_{|v|\geq T} \gamma_{ij}^{(v)} + \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T} \gamma_{ij}^{(v)} \right| < \frac{1}{T} \sum_{v=-q}^{q} |v| \left| \gamma_{ij}^{(v)} \right| + \frac{\varepsilon}{2}.$$

But the term $\sum_{v=-q}^{q} |v| \left| \gamma_{ij}^{(v)} \right|$ is a number that does not depend on T, so that $\frac{1}{T} \sum_{v=-q}^{q} |v| \left| \gamma_{ij}^{(v)} \right|$ can be made smaller than $\varepsilon/2$ for sufficiently large T. We thus conclude that, for any $\varepsilon > 0$, there is a T such that

$$\left| \sum_{|v|\geq T} \gamma_{ij}^{(v)} + \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T} \gamma_{ij}^{(v)} \right| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves that

$$\lim_{T\to\infty} TE\left[\left(\bar{X}_T - \mu \right) \left(\bar{X}_T - \mu \right)' \right] = \sum_{j=-\infty}^{\infty} \Gamma_j.$$

The above algebra is ~~tedious and boring~~ important as it shows the formula for the asymptotic covariance of the sample mean of a vector process. This is used, for example, for computing asymptotics of GMM with time dependent data or for computing robust standard errors of many estimators (e.g. OLS estimators where residuals could have autocorrelation and/or heteroskedasticity).

We now recall the two fundamental limit theorems: suppose that $X_1, X_2, ...$ are i.i.d. random variables with $E(X_t) = \mu$ and $E(X_t - \mu)^2 = \sigma^2 < \infty$. Then,

Law of large numbers (LLN): $\frac{1}{T}\sum_{t=1}^{T} X_t \to \mu$ (converges in probability, a.s., in L^2)

Central limit theorem (CLT): $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu\right) \Rightarrow N(0, \sigma^2)$ (converges in distribution)

The above results use independence and gives an idea of how quickly and in what sense the sample average converges to the population mean. In time series we typically don't have independence. There are, however, versions of the above theorems for dependent data. Let's consider how we prove the (weak) law of large number. For this we need the following results:

Markov inequality: Let $\phi(x) \geq 0$ be a non-decreasing function on R_+. Then, for any random variable $X \geq 0$ and constant $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[\phi(X)]}{\phi(a)}.$$

Proof:

$$\begin{aligned} E[\phi(X)] &= \Pr(\phi(X) \geq \phi(a)) E[\phi(X) | \phi(X) \geq \phi(a)] + \Pr(\phi(X) < \phi(a)) E[\phi(X) | \phi(X) < \phi(a)] \\ &\geq \Pr(\phi(X) \geq \phi(a)) E[\phi(X) | \phi(X) \geq \phi(a)] \\ &\geq \Pr(\phi(X) \geq \phi(a)) \phi(a) \end{aligned}$$

where the first inequality uses that $E[\phi(X) | \phi(X) < \phi(a)] \geq 0$ and the second inequality uses $E[\phi(X) | \phi(X) \geq \phi(a)] \geq \phi(a)$. Therefore,

$$\Pr(\phi(X) \geq \phi(a)) \leq \frac{E[\phi(X)]}{\phi(a)}.$$

To finish the proof, note the following inclusion of events

$$\{\omega \in \Omega : X(\omega) \geq a\} \subseteq \{\omega \in \Omega : \phi(X(\omega)) \geq \phi(a)\}.$$

If $\phi(x)$ is strictly increasing, the two events are equal, but if $\phi(x)$ is constant over some range, the inclusion can be strict. Therefore, the set inclusion implies

$$\Pr(X \geq a) \leq \Pr(\phi(X) \geq \phi(a)) \leq \frac{E[\phi(X)]}{\phi(a)}. \square$$

Chebyshev inequality: For any random variable X and constant $a > 0$,

$$\Pr\left(|X - E[X]| \geq a\right) \leq \frac{Var(X)}{a^2}.$$

Proof: Let $Z = |X - E[X]|$ and $\phi(z) = z^2$. Now apply Markov's inequality.□

We now provide a proof of the weak law of large numbers with iid data.

Theorem (WLLN): Let $X_1, X_2, ...$ be *i.i.d.* random variables with $E[X_t] = \mu$ and uniformly bounded variance $E\left[(X_t - \mu)^2\right] \leq B < \infty$. Let $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$. Then, for any $\varepsilon > 0$,

$$\lim_{T\to\infty} \Pr\left[\left|\bar{X}_T - \mu\right| > \varepsilon\right] = 0.$$

Proof: Note that $E\left[\bar{X}_T\right] = \mu$ and

$$\begin{aligned} Var\left(\bar{X}_T\right) &= Var\left(\frac{1}{T}\sum_{t=1}^{T} X_t\right) = \frac{1}{T^2} Var\left(\sum_{t=1}^{T} X_t\right) \\ &= \frac{1}{T^2}\sum_{t=1}^{T} Var(X_t) \leq \frac{TB}{T^2} = \frac{B}{T} \end{aligned}$$

where the third equality uses that X_t is iid and the inequality uses that the variance is bounded. Therefore, for any $\varepsilon > 0$, Chebyshev's inequality implies

$$\lim_{T\to\infty} \Pr\left[\left|\bar{X}_T - \mu\right| > \varepsilon\right] \leq \lim_{T\to\infty} \frac{Var\left(\bar{X}_T\right)}{\varepsilon^2} \leq \lim_{T\to\infty} \frac{B}{\varepsilon^2 T} = 0.\square$$

Note that we used that X_t is iid in two parts. First, for using that $E[X_t] = \mu$ for all t and, more importantly, for writing $Var\left(\sum_{t=1}^{T} X_t\right) = \sum_{t=1}^{T} Var(X_t)$.

What happens if the random variables are not i.i.d.? The variance of $\bar{X}_T$ is given by

$$Var\left(\frac{1}{T}\sum_{t=1}^{T} X_t\right) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(X_t, X_s).$$

A WLLN can be proved for covariance stationary processes by imposing restrictions on the autocovariances. In particular, suppose that $\{X_t\}$ is a covariance stationary process, so that $E(X_t) = \mu$ for all t and $Cov(X_s, X_t) = \gamma_{|s-t|}$ for all s, t with absolutely summable autocovariances, so that

$$\sum_{j=-\infty}^{\infty} |\gamma_j| = c < \infty.$$

Then,

$$\begin{aligned} T^2 Var\left(\bar{X}_T\right) &= \sum_{t=1}^{T}\sum_{s=1}^{T} Cov\left(X_t, X_s\right) = \sum_{t=1}^{T}\sum_{s=1}^{T} \gamma_{|t-s|} \\ &\leq \sum_{t=1}^{T}\sum_{s=1}^{T} \left|\gamma_{|t-s|}\right| \\ &\leq \sum_{t=1}^{T}\sum_{s=-\infty}^{\infty} \left|\gamma_{|t-s|}\right| \\ &\leq Tc \end{aligned}$$

where the last inequality uses absolute summability of autocovariances. It then follows that

$$Var\left(\bar{X}_T\right) \leq \frac{c}{T},$$

which is then used to prove the WLLN for dependent random variables under the assumption of stationarity and absolute summability of autocovariances.

A similar reasoning can be used to argue that the variance of the central limit theorem should change. Indeed, if the autocovariance function is absolutely summable, we showed above for the vector case that

$$\lim_{T\to\infty} T\ Var\left(\bar{X}_T\right) = \lim_{T\to\infty} Var\left(\sqrt{T}\ \bar{X}_T\right) = \sum_{j=-\infty}^{\infty} \gamma_j$$

where, $\gamma_j = Cov\left(X_t, X_{t-j}\right)$. This implies that, under an appropriate CLT,

$$\sqrt{T}\ \bar{X}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t \to N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right)$$

After reading the notes about time series in the frequency domain, please note that the asymptotic covariance of the sample mean is the spectral density of the process evaluated at frequency zero.

References:

[1] Cochrane, John (2005). Time Series for Macroeconomics and Finance. Manuscript.

[2] Gabel, R. A. and R. A. Roberts (1986). *Signals and Linear Systems*, 3rd Edition.

[3] Hamilton, James (1994). *Time Series Analysis*. Chapters 2 and 3

[4] Hansen, L. and T. Sargent (1991). Lecture notes on least squares predictions theory. In Rational Expectations Econometrics (Underground Classics in Economics). These notes provides detailed proofs with all the technicalities of most of the stuff we use. It can be found online here: http://www.tomsargent.com/books/TOMchpt.2.pdf

[5] Fuller, Wayne. (1995). *Introduction to Statistical Time Series*. Second Edition. Wiley Series in Probability and Statistics.

[6] Sargent, Thomas (1987). *Macroeconomic Theory*. Chapters 10 and 11.

Detalles del algoritmo RWZ (algorithm 2 en las slides)

This note fills in some of the details of the efficient algorithm in RWZ. We use the notation that we have in the slides

Let n be the number of variables.

- Short run restrictions are zero restrictions on S. Define

$$L_0 \equiv S$$

- Long-run restrictions are zeros imposed on IR_∞. Define

$$L_\infty = (1 - \mathbf{D})^{-1} S$$

  where $\mathbf{D} = \sum_{j=1}^{p} D_j$.

- We can embed short and long run restrictions in

$$f(S, D) = \begin{bmatrix} L_0 \\ L_\infty \end{bmatrix}$$

  Let $F = f(S, D)$. Note that F is a $2n \times n$ matrix.

- The identification constraints can be written as

$$Q_i F e_i = \mathbf{0}_{n \times 1}.$$

Let's dig into the algorithm.

1. Assume that the model is exactly identified
2. Let F be the matrix with short and long-run impact matrices. Let $Q_1, Q_2, ..., Q_n$ represent the identifying restrictions and assume that we have ordered the shocks so that $q_i = n - i$ for $i = 1, 2, ..., n$ is the rank of Q_i.

3. Perform a Cholesky decomposition of the covariance matrix of reduced form residuals,

$$V = chol\left(\Omega\right)$$

where $VV' = \Omega$.

4. Construct initial short and long run response matrices

$$\begin{aligned} L_0^* &= V \\ L_\infty^* &= \left(I - \mathbf{D}\right)^{-1} V \end{aligned}$$

And let

$$F^* = \left[\begin{array}{c} L_0^* \\ L_\infty^* \end{array} \right]$$

Of course, F^* need not satisfy the identifying assumptions.

5. By Theorem 1, there is a rotation matrix P such that $L_0 = L^* P$ and $L_\infty = L_\infty^* P$. Equivalently,

$$F = F^* P$$

satisfy the identifying assumptions.

6. RWZ provide an algorithm to construct P.

Set $j = 1$. Then, according to the algorithm,

$$\tilde{Q}_1 = Q_1 F^*.$$

Note that Q_1 has rank $n-1$, so that $Q_1 F^*$ has rank$< n$, $rank\left(\tilde{Q}_1\right) < n$. This implies that there exist an infinite number of non-trivial solutions to the system of equations $\tilde{Q}_1 \mathbf{x} = \mathbf{0}$. Then, there exist a unit-length vector $\mathbf{p}_1 \neq 0$ such that

$$\tilde{Q}_1 \mathbf{p}_1 = \mathbf{0}.$$

Note that

$$Q_1 F^* \mathbf{p}_1 = \mathbf{0}. \tag{1}$$

Set $j = 2$ and let

$$\underbrace{\tilde{Q}_2}_{n+1\times n} = \left[\begin{array}{c} Q_2F^* \\ \mathbf{p}_1' \end{array} \right]$$

Note that $\tilde{Q}_2$ is of size $n+1 \times n$. Moreover, since $rank\,(Q_2) = n-2$, $rank\,(Q_2F^*) \leq n-2$. Therefore, $rank(\tilde{Q}_2) < n$. Hence, there exist a unit-length vector $\mathbf{p}_2$ such that

$$\tilde{Q}_2\mathbf{p}_2 = \mathbf{0}_{n+1\times 1}$$

This implies two things:

$$\begin{aligned} Q_2F^*\mathbf{p}_2 &= \mathbf{0}_{n\times 1} \\ \mathbf{p}_1'\mathbf{p}_2 &= 0. \end{aligned} \tag{2}$$

Set $j = 3$ and let

$$\underbrace{\tilde{Q}_3}_{n+2\times n} = \left[\begin{array}{c} Q_3F^* \\ \mathbf{p}_1' \\ \mathbf{p}_2' \end{array} \right].$$

Now, since $rank\,(Q_3) = n-3$, then $rank\,(Q_3F^*) \leq n-3$ and, hence, $rank(\tilde{Q}_3) < n$. Therefore, there exist a unit-length vector $\mathbf{p}_3$ such that

$$\tilde{Q}_3\mathbf{p}_3 = \mathbf{0}_{n+2\times 1}.$$

This implies:

$$\begin{aligned} Q_3F^*\mathbf{p}_3 &= \mathbf{0}_{n\times 1} \\ \mathbf{p}_1'\mathbf{p}_3 &= 0 \\ \mathbf{p}_2'\mathbf{p}_3 &= 0. \end{aligned} \tag{3}$$

Continue doing this n times. In this way, we are constructing a matrix with columns $\mathbf{p}_1$, $\mathbf{p}_2$, ...,

$\mathbf{p}_n$ such that

$$\begin{aligned} Q_j F^* \mathbf{p}_j &= \mathbf{0}_{n\times 1} \\ \mathbf{p}_j' \mathbf{p}_j &= 1 \\ \mathbf{p}_j' \mathbf{p}_k &= 0 \end{aligned} \tag{4}$$

Let

$$P = \left[\begin{array}{ccccc} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & ... & \mathbf{p}_n \end{array} \right]$$

so that

$$PP' = I.$$

Now let

$$F = F^* P.$$

And let's check that the identifying assumptions are satisfied for the matrix F. We must have

$$Q_j F e_j = 0_{n\times 1}$$

for $j = 1, 2, ..., n$. But

$$\begin{aligned} Q_j F e_j &= Q_j F^* P e_j \\ &= Q_j F^* \left[\begin{array}{ccccc} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & ... & \mathbf{p}_n \end{array} \right] \times \left[\begin{array}{c} 0 \\ \vdots \\ 1 \text{ (row } j\text{)} \\ \vdots \\ 0 \end{array} \right] \\ &= Q_j F^* \mathbf{p}_j \\ &= \mathbf{0} \end{aligned}$$

where the last equality follows from equation (4). Since this holds for all j, we have

$$Q_j F e_j = \mathbf{0}$$

for $j = 1, 2, ..., n$ once we set $F = F^* P$ with P constructed as above. This gives the desired rotation matrix P and we set $S = VP$.

– Lecture Notes 1

An introduction to Bayesian Vector Autoregressions

These lecture notes contain derivations used in the BVAR set of slides and a discussion of certain aspects of Bayesian estimation. We first define some useful probability distributions that will be used throughout the notes. After these initial definitions, we begin with the Bayesian analysis of the linear regression model. Then, we discuss more general Bayesian VARs.

These notes are not a complete treatment of Bayesian econometrics. If you are interested in this topic, you should consult some of the many books on Bayesian econometrics.

Preliminaries

In this section, we define the multivariate normal density, the Wishart density, and the inverse Wishart density. We also discuss Bayes's Theorem, which is the most important tool in Bayesian analysis. Throughout this note, we denote the determinant of a matrix A by $|A|$ and the trace of a square matrix B by $\text{tr}(B)$.[1] We also use $p(\cdot)$ to denote the probability of an event or the probability density function of a random variable or vector, depending on the context.

To warm up, we start by defining the univariate normal and the Gamma distributions. The Gamma distribution will also be used when we consider the Bayesian analysis of the simple linear regression model.

Definition 1.1 — Normal random variable

Let X be a continuous random variable defined for all real numbers. We say that X has a **normal distribution** with mean $\mu \in R$ and variance $\sigma^2 \in R_+$, denoted by $X \sim N(\mu, \sigma^2)$, if its probability density function is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}. \tag{1.1}$$

When $\mu = 0$ and $\sigma^2 = 1$ we obtain the standard normal distribution.

With normal random variables, we can construct other random variables of interest. For example, if we consider the sum of squares of n independent standard normal variables, we

[1] Recall that the trace of a square matrix is defined as the sum of the diagonal elements.

obtain the chi-square distribution, which is often used in the estimation of variance and hypothesis testing. In particular, if Y_i for $i = 1, \ldots, n$ are independent standard normal random variables (so that $Y_i \sim N(0, 1)$), the random variable $X = Y_1^2 + Y_2^2 + ...Y_n^2$ has a chi-square distribution with n degrees of freedom, often written as $X \sim \chi_n^2$. The chi-square distribution is a special case of the Gamma distribution.

The Gamma distribution is widely used in Bayesian analysis as a prior distribution for parameters that must be positive, such as variances. The Gamma distribution depends on two parameters, a shape parameter k and a rate parameter θ. Above, we mentioned that the sum of squares of independent standard normal random variables is distributed as a chi-square. For the case of arbitrary variances, the sum of squares now follows a Gamma distribution. In particular, suppose that $Y_i \sim N(0, \sigma^2)$ for $i = 1, 2, \ldots, n$ are independent normal random variables. Then, $X = Y_1^2 + Y_2^2 + ...Y_n^2$ is distributed as a Gamma random variable with shape parameter $k = n/2$ and rate parameter $\theta = 1/(2\sigma^2)$.

The Gamma distribution is defined as follows:

Definition 1.2 — Gamma distribution

Let X be a continuous random variable defined for $X \geq 0$. We say that X has a **Gamma** distribution with shape parameter $k > 0$ and rate parameter $\theta > 0$, denoted by $X \sim Gamma(k, \theta)$, if its probability density function is

$$f(x|k, \theta) = \frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-\theta x}, \quad x \geq 0, \tag{1.2}$$

where $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t}\, dt$ is the Gamma function.

We now consider multivariate generalizations of the previous formulas. We start with the multivariate normal distribution and next define the Wishart distribution, which is a generalization of the Gamma distribution to the multivariate case. It will also become useful below to define the inverse Wishart distribution, which is the distribution of the inverse of a Wishart.

Definition 1.3 — Multivariate Normal distribution

Let $\boldsymbol{X} \in R^k$ be a random vector with k elements, $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)'$, and let $\boldsymbol{\mu} \in R^k$, and Σ be a positive definite symmetric $k \times k$ matrix. We say that $\boldsymbol{X}$ is normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix Σ, denoted by $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, if the probability density function of $\boldsymbol{X}$ is given by

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right) \tag{1.3}$$

We now consider the multivariate generalization of the sum of squares of mean-zero normal random variables. Now we have to consider all possible cross products between the random variables.

Definition 1.4 — Wishart distribution

Let $X_i \in R^l$ for $i = 1, 2, \ldots, m$ be an $l-$dimensional i.i.d. random vector distributed according to $X_i \sim N(\mathbf{0}_{l\times 1}, \Sigma)$, where Σ is an $l \times l$ positive definite matrix. Let $Z = \sum_{i=1}^{m} X_i X_i'$, which is an $l \times l$ random matrix with the sum of all squared terms in X_i. The distribution of Z is called a ($l-$dimensional) **Wishart distribution** with scale matrix Σ and m degrees of freedom. The probability density function of Z is given by

$$f(Z|\Sigma, m) = \frac{|Z|^{(m-l-1)/2}}{2^{ml/2}\,|\Sigma|^{m/2}\,\Gamma_l(m/2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}Z)\right] \tag{1.4}$$

where $\Gamma_l(t) = \pi^{l(l-1)/4}\prod_{i=1}^{l}\Gamma(t-(i-1)/2)$ is the multivariate Gamma function. We write $Z \sim W_l(\Sigma, m)$

If the random matrix Z follows a Wishart distribution, Z^{-1} follows an Inverse Wishart distribution:

Definition 1.5 — Inverse Wishart distribution

We say that Z follows an **inverse Wishart distribution** with scale matrix Ψ and m degrees of freedom, denoted by $Z \sim iW(\Psi, m)$, if Z^{-1} has a Wishart distribution with scale matrix Ψ^{-1} and m degrees of freedom, so that $Z^{-1} \sim W(\Psi^{-1}, m)$. The probability density function of the inverse Wishart is given by

$$f(Z|\Psi, m) = \frac{|\Psi|^{m/2}\,|Z|^{(m+l+1)/2}}{2^{ml/2}\,|\Sigma|^{m/2}\,\Gamma_l(m/2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Psi Z^{-1})\right]. \tag{1.5}$$

For completeness, we also state the Bayes Theorem:

Theorem 1.1 — Bayes' rule

Let x and y be random variables (or vectors). Then

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\int p(x|y)p(y)dy}. \tag{1.6}$$

Proof. The proof follows from the definition of conditional probabilities:

$$p(x|y) = \frac{p(x,y)}{p(y)} \Rightarrow p(x,y) = p(x|y)p(y),$$
$$p(y|x) = \frac{p(x,y)}{p(x)} \Rightarrow p(x,y) = p(y|x)p(x).$$

Equating both terms and rearranging gives the first equality in equation (1.6). To obtain the second equality use the Law of total probability: $\int p(x|y)p(y)dy = \int p(x,y)dy = p(x)$. ■

Bayesian analysis of the linear regression model

Consider the standard linear regression model

$$y_i = x_i'\beta + \epsilon_i \tag{1.7}$$

where $i = 1, 2, ..., n$ is the number of observations, y_i is the dependent variable, x_i is a vector with k regressors, $\beta \in R^k$ and $\epsilon_i \sim N(0, \sigma^2)$.

Stacking the observations in (1.7) we obtain

$$Y = X\beta + \epsilon; \qquad \epsilon \sim N(0, \sigma^2 I_n), \tag{1.8}$$

where Y is $n \times 1$, X is $n \times k$ and ϵ is $n \times 1$.

Since $\epsilon = Y - X\beta$ is distributed as a multivariate normal, the likelihood function of (1.8) is

$$\begin{aligned} p(Y|\beta, \sigma^2) =& (2\pi)^{-\frac{n}{2}} \left|\sigma^2 I_n\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(Y - X\beta)'\left(\sigma^2 I_n\right)^{-1}(Y - X\beta)\right) \\ =& (2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right). \end{aligned} \tag{1.9}$$

Consider first the Maximum Likelihood Estimator (MLE) of β. To that end, take logs on both sides of (1.9) and obtain the log-likelihood function

$$\begin{aligned} l(Y|\beta, \sigma^2) =& -\frac{n}{2}\log(2\pi) + \log\left((\sigma^2)^{-\frac{n}{2}}\right) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \\ =& -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left[Y'Y - 2\beta'X'Y + \beta'X'X\beta\right]. \end{aligned}$$

Take the first order condition with respect to β and rearrange to obtain

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

That is, the MLE estimator is identical to the OLS estimator of β.

We can write

$$\hat{\beta} = (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\beta + \epsilon) = \beta + (X'X)^{-1} X'\epsilon.$$

and note that

$$E\left[\hat{\beta}\right] = \beta$$

and

$$\begin{aligned} E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right] =& E\left[(X'X)^{-1} X'\epsilon\epsilon' X\left((X'X)^{-1}\right)'\right] \\ =& (X'X)^{-1} X' E\left[\epsilon\epsilon'\right] X\left((X'X)^{-1}\right)' \\ =& (X'X)^{-1} X'\sigma^2 I_T X\left((X'X)^{-1}\right)' \\ =& \sigma^2 (X'X)^{-1}(X'X)\left((X'X)^{-1}\right)' \\ =& \sigma^2\left((X'X)'\right)^{-1} \\ =& \sigma^2 (X'X)^{-1}. \end{aligned}$$

It then follows that

$$\hat{\beta} \sim N\left(\beta, \sigma^2 \left(X'X\right)^{-1}\right).$$

The MLE relies solely on data for parameter estimation. On the other hand, Bayesian analysis offers the ability to incorporate prior beliefs about the parameters through prior density distributions for β and σ^2. In our case, we aim to perform a Bayesian analysis where we treat β and σ^2 as random variables. Specifically, our objective is to determine the posterior distribution of β and σ^2 after observing the data Y. To achieve this, we combine prior distributions of β and σ^2 with the information contained in the data using the likelihood function $p(Y|\beta, \sigma^2)$. The result is the posterior distribution $p(\beta, \sigma^2|Y)$.

However, instead of directly obtaining the joint posterior $p(\beta, \sigma|Y)$, we will compute the conditional posterior probabilities $p(\beta|\sigma^2, Y)$ and $p(\sigma^2|\beta, Y)$. In this example, we introduce the **Gibbs Sampler**, a Markov Chain Monte Carlo sampling technique used to draw samples from the joint posterior $p(\beta, \sigma|Y)$. It relies on the knowledge of the conditional posteriors $p(\beta|\sigma^2, Y)$ and $p(\sigma^2|\beta, Y)$, which we will derive. First, we consider the case where σ^2 is known, and derive a closed form expression of the conditional posterior $p(\beta|\sigma^2, y)$. Then, assuming knowledge of β, we derive the closed form expression of the conditional posterior $p(\sigma^2|\beta, Y)$. Finally, we explain how to use the Gibbs Sampler to draw samples from the joint posterior $p(\beta, \sigma^2|Y)$.

Bayesian analysis with unknown β and known σ^2

Let's pretend for the moment that we know the variance σ^2 of the error term ϵ_i. This is equivalent to saying that we are conditioning on a given value of σ^2.

We start with a prior belief about β in the form of a distribution $p(\beta)$, which we assume to be normal:

$$\beta \sim N(\beta_0, \Sigma_0),$$

where β_0 is $k \times 1$ and Σ_0 is $k \times k$ —there are k regressors in the linear model (1.7).

Next, we construct the likelihood function $p(Y|\beta, \sigma^2)$ as above. Finally, we combine the prior belief with the likelihood function to obtain the posterior distribution of the parameters $p(\beta|\sigma^2, Y)$ given the data.

Bayes's theorem then implies

$$\begin{aligned} p(\beta|Y, \sigma^2) =& \frac{p(Y|\beta, \sigma^2)p(\beta)}{\int_\beta p(Y|\beta, \sigma^2)p(\beta)d\beta} \\ \propto& p(Y|\beta, \sigma^2)p(\beta), \end{aligned}$$

where the symbol $\propto$ means "is proportional to".

The prior density of the parameter β is given by

$$\begin{aligned} p(\beta) =& (2\pi)^{-\frac{k}{2}} |\Sigma_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\beta - \beta_0\right)' \Sigma_0^{-1} \left(\beta - \beta_0\right)\right) \\ \propto& \exp\left[-\frac{1}{2}\left(\beta - \beta_0\right)' \Sigma_0^{-1} \left(\beta - \beta_0\right)\right], \end{aligned}$$

Since we assume that σ^2 is known, we can write the likelihood function (1.9) as proportional to

$$p\left(Y|\beta,\sigma^2\right) \propto \exp\left(-\frac{1}{2\sigma^2}\left(Y-X\beta\right)'\left(Y-X\beta\right)\right)$$

because $(\sigma^2)^{n/2}$ is a constant that we assume known and, therefore, we can ignore.

Combining the prior with the likelihood function, we obtain the kernel of the posterior distribution:

$$\begin{aligned}
p(\beta|Y,\sigma^2) \propto & p(Y|\beta,\sigma^2)p(\beta) \\
& \propto \exp\left[-\frac{1}{2\sigma^2}\left(Y-X\beta\right)'\left(Y-X\beta\right)\right]\exp\left[-\frac{1}{2}\left(\beta-\beta_0\right)'\Sigma_0^{-1}\left(\beta-\beta_0\right)\right] \\
& \propto \exp\left[-\frac{1}{2}\left(\beta-\beta_0\right)'\Sigma_0^{-1}\left(\beta-\beta_0\right)-\frac{1}{2\sigma^2}\left(Y-X\beta\right)'\left(Y-X\beta\right)\right] \\
& \propto \exp\left[-\frac{1}{2}\left(\beta-\beta_1\right)'\Sigma_1^{-1}\left(\beta-\beta_1\right)\right],
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_1^{-1} =& \left(\Sigma_0^{-1}+\frac{1}{\sigma^2}X'X\right) \\
\beta_1 =& \Sigma_1\left(\Sigma_0^{-1}\beta_0+\frac{1}{\sigma^2}X'Y\right).
\end{aligned}$$

Now, since the previous equation is the kernel of a normal distribution, we know that the conditional posterior is normally distributed, so that

$$\beta|\{Y,\sigma^2\} \sim N\left(\beta_1,\Sigma_1\right). \tag{1.10}$$

That is, the conditional posterior $p(\beta|Y,\sigma^2)$ is (multivariate) normally distributed with mean β_1 and variance Σ_1.

We now prove the previous claim. Consider the term inside the exponential function

$$Z = -\frac{1}{2}\left(\beta-\beta_0\right)'\Sigma_0^{-1}\left(\beta-\beta_0\right)-\frac{1}{2\sigma^2}\left(Y-X\beta\right)'\left(Y-X\beta\right).$$

We want to write this expression in the form

$$Z = -\frac{1}{2}\left(\beta-\beta_1\right)'\Sigma_1^{-1}\left(\beta-\beta_1\right)+\text{a constant},$$

for some β_1 and Σ_1 that are yet to be found. This trick is called "completing the square" and is used a lot. First note that

$$\left(\beta-\beta_1\right)'\Sigma_1^{-1}\left(\beta-\beta_1\right) = \beta'\Sigma_1^{-1}\beta-2\beta_1'\Sigma_1^{-1}\beta+\beta_1'\Sigma_1^{-1}\beta_1. \tag{1.11}$$

Now we work with the term inside the exponent of the posterior kernel

$$\begin{aligned}
& \left(\beta-\beta_0\right)'\Sigma_0^{-1}\left(\beta-\beta_0\right)+\frac{1}{\sigma^2}\left(Y-X\beta\right)'\left(Y-X\beta\right) \\
& = \beta'\Sigma_0^{-1}\beta-2\beta_0'\Sigma_0^{-1}\beta+\beta_0'\Sigma_0^{-1}\beta_0+\frac{1}{\sigma^2}Y'Y-\frac{2}{\sigma^2}Y'X\beta+\frac{1}{\sigma^2}\beta'X'X\beta \\
& = \beta'\left(\Sigma_0^{-1}+\frac{1}{\sigma^2}X'X\right)\beta-2\left(\beta_0'\Sigma_0^{-1}+\frac{2}{\sigma^2}Y'X\right)\beta+\beta_0'\Sigma_0^{-1}\beta_0+\frac{1}{\sigma^2}Y'Y.
\end{aligned} \tag{1.12}$$

Compare this last expression with the right side of equation (1.11). They are quite similar. Let's define

$$\Sigma_1^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2} X'X$$

and

$$\beta_1'\Sigma_1^{-1} = \beta_0'\Sigma_0^{-1} + \frac{2}{\sigma^2} Y'X \Rightarrow \beta_1 = \Sigma_1 \left(\Sigma_0^{-1}\beta_0 + \frac{2}{\sigma^2} X'Y \right).$$

With these definitions, equation (1.12) can be written as

$$\begin{aligned}
&\beta' \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X \right) \beta - 2 \left(\beta_0'\Sigma_0^{-1} + \frac{2}{\sigma^2} Y'X \right) \beta + \beta_0'\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2} Y'Y \\
=&\beta'\Sigma_1^{-1}\beta - 2\beta_1'\Sigma_1^{-1}\beta + \beta_1'\Sigma_1^{-1}\beta_1 - \beta_1'\Sigma_1^{-1}\beta_1 + \beta_0'\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2} Y'Y \\
=&\left(\beta - \beta_1\right)' \Sigma_1^{-1} \left(\beta - \beta_1\right) + \beta_0'\Sigma_0^{-1}\beta_0 - \beta_1'\Sigma_1^{-1}\beta_1 + \frac{1}{\sigma^2} Y'Y \\
=&\left(\beta - \beta_1\right)' \Sigma_1^{-1} \left(\beta - \beta_1\right) + C
\end{aligned}$$

where C is a constant (i.e. does not depend on the random variable β). This term is like (1.11) for the proposed β_1 and Σ_1 with a constant in the exponent. Also, note that in the second line we added and subtracted the term $\beta_1'\Sigma_1^{-1}\beta_1$.

Summarizing, we showed that the kernel of the posterior is

$$p(\beta|Y, \sigma^2) \propto \exp\left[-\frac{1}{2} \left(\beta - \beta_1\right)' \Sigma_1^{-1} \left(\beta - \beta_1\right) \right]$$

which proves that the posterior is normal $\beta|\left\{Y, \sigma^2\right\} \sim N\left(\beta_1, \Sigma_1\right)$ with

$$\begin{aligned}
\Sigma_1^{-1} =& \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X \right) \\
\beta_1 =& \Sigma_1 \left(\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2} X'Y \right).
\end{aligned}$$

But also note that the OLS estimator is $\hat{\beta} = \left(X'X\right)^{-1} X'Y \Rightarrow X'Y = \left(X'X\right)\hat{\beta}$. Therefore,

$$\begin{aligned}
\beta_1 =& \Sigma_1 \left(\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2} \left(X'X\right)\hat{\beta} \right) \\
=& \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X \right)^{-1} \left(\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2} \left(X'X\right)\hat{\beta} \right) \\
=& W \times \beta_0 + \left(I - W\right) \times \hat{\beta},
\end{aligned}$$

where $W = \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X \right)^{-1} \Sigma_0^{-1}$.

That is, the mean of the posterior distribution of β is a weighted average of the prior mean, β_0, and of the OLS estimator $\hat{\beta}$. Note that if the variance of the prior goes to infinity (e.g. $\Sigma_0 = m\tilde{\Sigma}_0$ with $m \to \infty$, then $\Sigma_0^{-1} \to 0$) we get that $W \to \mathbf{0}$ and the mean of the posterior is the OLS estimate $\beta_1 = \hat{\beta}$. On the other hand, if the data is uninformative, so that $\sigma \to \infty$, we get $W \to I$, $\beta_1 \to \beta_0$ and the mean of the posterior is the mean of the prior.

In addition, note that setting $\Sigma_0^{-1} = \frac{\lambda}{\sigma^2} I$ and $\beta_0 = 0$ gives the Ridge regression:[2]

$$\begin{aligned}\beta_1 &= (\lambda I + X'X)^{-1} (X'X)\, \hat{\beta} \\ &= (\lambda I + X'X)^{-1} X'Y.\end{aligned}$$

Example 1: Suppose that the linear regression is

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim N(0, 1).$$

Simulate data assuming $\beta_0 = -1$, $\beta_1 = 2$, $n = 3$ and $x_i \sim \text{Uniform}(0, 1)$. Note that the variance of ϵ_i is fixed at 1. Consider two priors for β:

a) Uniform prior $p(\beta) \propto 1$.

b) Normal prior $p(\beta) \sim N(0, I_2)$.

See the slides for the posteriors in this example.

Exercise: Derive the posterior for the uniform prior for β.

Bayesian analysis with unknown σ^2 and known β

We now reverse the roles and assume that β is known and the variance parameter σ^2 is unknown. We aim to estimate the conditional posterior of σ^2. A typical prior for the (inverse of the) variance of a linear regression is the Gamma distribution (1.2) that we discussed in the Preliminaries section.

Imposing a prior to the variance σ^2 is equivalent to imposing a prior to the precision parameter $\tau \equiv 1/\sigma^2$. We will assume that prior of the precision is a Gamma distribution $\tau \sim Gamma(a_0, b_0)$. The prior distribution of τ is thus proportional to

$$p(\tau) \propto \tau^{a_0 - 1} \exp(-\tau b_0).$$

The mean of τ is $E(\tau) = a_0/b_0$ and the variance of τ is $\text{Var}(\tau) = a_0/b_0^2$. The hyperparameters a_0 and b_0 determine the shape of the Gamma distribution.

Using τ instead of σ^2, we can rewrite the kernel of the likelihood function (1.9) as

$$p(Y|\beta, \tau) \propto \tau^{\frac{n}{2}} \exp\left[-\frac{\tau}{2} (Y - X\beta)' (Y - X\beta)\right]$$

If we combine the kernel of the prior with the kernel of the likelihood, we obtain the kernel of the posterior distribution for τ:

$$\begin{aligned}p(\tau|\beta, Y) \propto & p(Y|\tau, \beta) p(\tau) \\ \propto & \tau^{\frac{n}{2}} \tau^{a_0 - 1} \exp[-b_0 \tau] \exp\left[-\frac{\tau}{2} (Y - X\beta)' (Y - X\beta)\right] \\ \propto & \tau^{a_0 + \frac{n}{2} - 1} \exp\left[-\tau \left(b_0 + \frac{1}{2} (Y - X\beta)' (Y - X\beta)\right)\right]\end{aligned}$$

[2] If you don't know what a Ridge regression is, don't worry, you can ignore this paragraph.

Note that this is the kernel of a Gamma distribution

$$\tau|\{\beta, Y\} \sim Gamma(a_1, b_1) \tag{1.13}$$

with parameters

$$\begin{aligned} a_1 =& a_0 + \frac{n}{2}, \\ b_1 =& b_0 + \frac{1}{2}(Y - X\beta)'(Y - X\beta). \end{aligned}$$

So if we know β and the prior for the precision is a Gamma distribution, we know that the conditional posterior of the precision is also a Gamma distribution but with different hyperparameters.

Bayesian analysis for unknown β and σ^2

Given a normal prior for β and a Gamma prior for $\tau = 1/\sigma^2$, we were able to compute the conditional posteriors $p(\beta|\tau, Y)$ and $p(\tau|\beta, Y)$. But we want to sample from the joint posterior distribution $p(\beta, \tau|Y)$. How do we do it?

In this case there is a powerful tool to draw samples from $p(\beta, \tau|Y)$ given that we know the two conditional posteriors $p(\beta|\tau, Y)$ and $p(\tau|\beta, Y)$. This tool is the Gibbs Sampler. Specifically, we will construct a Markov chain $\{\beta^j, \tau^j\}$ for $j = 1, 2, \ldots, N$ for large N that will converge to the target posterior density $p(\beta, \tau|Y)$. We are not going to prove this result, but it can be shown that the Gibbs Sampler generates a Markov chain whose invariant distribution is precisely the posterior $p(\beta, \tau|Y)$. The algorithm is as follows:[3]

Algorithm: Gibbs Sampler for the linear regression model: `Choose a large` N`, an arbitrary` τ^0`, and set` $j = 1$`. Then iterate on the following loop:`

`a) Draw` β^j `from` $p(\beta|\tau^{j-1})$

`b) Draw` τ^j `from` $p(\tau|\beta^j)$

`c) Store` $\{\beta^j, \tau^j\}$`, set` $j = j + 1$`, and return to step a) while` $j < N$.

Bayesian Vector Autoregressions

Consider the vector autoregression (VAR)

$$Y_t = c + D_1 Y_{t-1} + D_2 Y_{t-2} + \ldots + D_p Y_{t-p} + v_t \tag{1.14}$$

where Y_t is $n \times 1$ (n time series), v_t is an $n \times 1$ vector of errors such that $v_t \sim N(0, \Omega)$, c is an $n \times 1$ vector, and D_j are $n \times n$ coefficient matrices. The (unrestricted) $VAR(p)$ is a multiple equation regression in which the regressors of each equation are the p lagged values of all the variables in Y. In particular, there are n equations, each of which has $k = np + 1$ regressors (the 1 is the constant) so there is a total of $kn = n^2p + 1$ coefficients in the constant and

[3] See the slides for the results.

the D_j matrices. We assume we have data for $t = 1, 2, ..., T$ taking the initial observations $y_{-p+1}, y_{-p+2}, ..., y_0$ as given. In other words, we have $T + p$ observations and let T be the number of effective observations for each variable in Y.

Transpose the VAR(p) (1.14) and write it as

$$\begin{aligned} Y_t' =& Y_{t-1}'D_1' + Y_{t-2}'D_2' + \ldots + Y_{t-p}'D_p' + c' + v_t' \\ =& \begin{bmatrix} Y_{t-1}' & Y_{t-2}' & \cdots & Y_{t-p}' & c' \end{bmatrix} \begin{bmatrix} D_1' \\ D_2' \\ \vdots \\ D_p' \\ c' \end{bmatrix} + v_t' \\ =& X_t\boldsymbol{\beta} + v_t', \end{aligned}$$

where $X_t = \begin{bmatrix} Y_{t-1}' & Y_{t-2}' & \cdots & Y_{t-p}' & 1 \end{bmatrix}$ is $1 \times (np+1)$ and

$$\underbrace{\boldsymbol{\beta}}_{(np+1)\times n} = \begin{bmatrix} D_1' \\ D_2' \\ \vdots \\ D_p' \\ c' \end{bmatrix}.$$

Let $k \equiv np + 1$, so that X_t is $1 \times k$ and $\boldsymbol{\beta}$ is $k \times n$ matrix. Note that

$$Y_t' = X_t\boldsymbol{\beta} + v_t' \tag{1.15}$$

is a system of n equations, one for each time period $t = 1, 2, ..., T$.

Now we stack vertically all the variables across observations. In particular, let

$$\boldsymbol{Y}_{T\times n} = \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_T' \end{bmatrix}; \quad \boldsymbol{X}_{T\times k} = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix}; \quad \boldsymbol{V}_{T\times n} = \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_T' \end{bmatrix}$$

Then we can write the system of equations simultaneously for all periods as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{V}. \tag{1.16}$$

Now vectorize the equations and note that $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}I_n \Rightarrow \operatorname{vec}(\boldsymbol{X}\boldsymbol{\beta}) = \operatorname{vec}(\boldsymbol{X}\boldsymbol{\beta}I_n) = (I_n \otimes \boldsymbol{X}) \operatorname{vec}(\boldsymbol{\beta})$. Then,

$$\operatorname{vec}(\boldsymbol{Y}) = (I_n \otimes \boldsymbol{X}) \operatorname{vec}(\boldsymbol{\beta}) + \operatorname{vec}(\boldsymbol{V}). \tag{1.17}$$

Let

$$y_{nT\times 1} = \operatorname{vec}(\boldsymbol{Y}); \ \beta_{kn\times 1} = \operatorname{vec}(\boldsymbol{\beta}); \text{ and } v_{nT\times 1} = \operatorname{vec}(\boldsymbol{V}) \sim N(0, \Omega \otimes I_T).$$

Then we can write the VAR(p) (1.14) in the format of a standard linear regression model for a single dependent variable:

$$y = (I_n \otimes \boldsymbol{X})\beta + v. \tag{1.18}$$

Exercise: Prove that $E(vv') = \Omega \otimes I_T$. (This is ugly but straightforward algebra).

We will now write the likelihood function in two different (but equivalent) ways that will prove useful later on. Since $v_{nT\times 1} \sim N(0, \Omega \otimes I_T)$, we know that

$$y \sim N\left(\left(I_n \otimes \boldsymbol{X}\right)\beta, \Omega \otimes I_T\right).$$

Hence, the likelihood function is

$$L(\beta, \Omega|Y) = \frac{|\Omega \otimes I_T|^{-1/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\left(y - \left(I_n \otimes \boldsymbol{X}\right)\beta\right)'\left(\Omega \otimes I_T\right)^{-1}\left(y - \left(I_n \otimes \boldsymbol{X}\right)\beta\right)\right]. \quad (1.19)$$

Now, use that for arbitrary matrices $A_{n\times n}$ and $B_{m\times m}$, $\det(A \otimes B) = \det(A)^m \det(B)^n$. Then we can write

$$|\Omega \otimes I_T| = |\Omega|^T.$$

Also, the property $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ implies $(\Omega \otimes I_T)^{-1} = \Omega^{-1} \otimes I_T^{-1}$. Using those two results, we write the likelihood as

$$L(\beta, \Omega|Y) = \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\left(y - \left(I_n \otimes \boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1} \otimes I_T\right)\left(y - \left(I_n \otimes \boldsymbol{X}\right)\beta\right)\right] \quad (1.20)$$

Now write the likelihood in a different, but equivalent, way that uses the matrix form (1.16). To that end, we use the following results about matrices, vectorization, and kronecker products: for comformable matrices B, C, D, we have

$$\begin{aligned} \operatorname{vec}(BCD) &= (D' \otimes B)\operatorname{vec}(C) && (1.21)\\ \operatorname{tr}(B'C) &= \operatorname{vec}(B')\operatorname{vec}(C) && (1.22)\\ \operatorname{tr}(BCD) &= \operatorname{tr}(CDB) = \operatorname{tr}(DBC) && (1.23) \end{aligned}$$

where $\operatorname{vec}(\cdot)$ is the vectorization operator and $\operatorname{tr}(\cdot)$ is the trace function.

Using that $y - (I_n \otimes \boldsymbol{X})\beta = v = \operatorname{vec}(\boldsymbol{V})$, we can write the quadratic form in the exponent of (1.20) as

$$\begin{aligned} \left(y - \left(I_n \otimes \boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1} \otimes I_T^{-1}\right)\left(y - \left(I_n \otimes \boldsymbol{X}\right)\beta\right) &= \operatorname{vec}(\boldsymbol{V})'\left(\Omega^{-1} \otimes I_T\right)\operatorname{vec}(\boldsymbol{V})\\ &= \operatorname{vec}(\boldsymbol{V})'\operatorname{vec}\left(I_T\boldsymbol{V}\Omega^{-1}\right)\\ &= \operatorname{vec}(\boldsymbol{V})'\operatorname{vec}\left(\boldsymbol{V}\Omega^{-1}\right)\\ &= \operatorname{tr}\left(\boldsymbol{V}'\boldsymbol{V}\Omega^{-1}\right)\\ &= \operatorname{tr}\left(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V}\right) \quad (1.24) \end{aligned}$$

where in the second equality we used (1.21), the fourth equality uses (1.22), and the last equality uses (1.23). Then we can equivalently write the likelihood function as

$$L(\beta, \Omega|Y) = \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V}\right)\right]. \quad (1.25)$$

MLE estimation of β

Rewrite the quadratic term in (1.20) as follows:

$$\begin{aligned}
&(y-(I_n\otimes \boldsymbol{X})\beta)'\left(\Omega^{-1}\otimes I_T\right)(y-(I_n\otimes \boldsymbol{X})\beta)\\
&=y'\left(\Omega^{-1}\otimes I_T\right)y-2\beta'(I_n\otimes \boldsymbol{X}')\left(\Omega^{-1}\otimes I_T\right)y+\beta'(I_n\otimes \boldsymbol{X}')\left(\Omega^{-1}\otimes I_T\right)(I_n\otimes \boldsymbol{X})\beta\\
&=y'\left(\Omega^{-1}\otimes I_T\right)y-2\beta'\left(\Omega^{-1}\otimes \boldsymbol{X}'\right)y+\beta'\left(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right)\beta
\end{aligned}$$

where the last equality uses

$$(I_n\otimes \boldsymbol{X}')\left(\Omega^{-1}\otimes I_T\right)=\left(\Omega^{-1}\otimes \boldsymbol{X}'\right)$$

and

$$(I_n\otimes \boldsymbol{X}')\left(\Omega^{-1}\otimes I_T\right)(I_n\otimes \boldsymbol{X})=\left(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right)$$

Taking logs in (1.20), we obtain the log-likelihood function

$$l(\beta,\Omega|Y)=-\frac{nT}{2}\log(2\pi)-\frac{T}{2}\log|\Omega|-\frac{1}{2}\left[y'\left(\Omega^{-1}\otimes I_T\right)y-2\beta'\left(\Omega^{-1}\otimes \boldsymbol{X}'\right)y+\beta'\left(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right)\beta\right].$$

To obtain the MLE estimator of β, take the first order condition of $l(\beta,\Omega|Y)$ and equate to zero:

$$\left(\Omega^{-1}\otimes \boldsymbol{X}'\right)y-\left(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right)\beta^{\text{mle}}=0. \tag{1.26}$$

Solving for β^{mle},

$$\beta^{\text{mle}}=\left(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right)^{-1}\left(\Omega^{-1}\otimes \boldsymbol{X}'\right)y.$$

But using the property $(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$, we have $(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X}))^{-1}=\Omega\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}$, so that

$$\beta^{\text{mle}}=\left(\Omega\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)\left(\Omega^{-1}\otimes \boldsymbol{X}'\right)y.$$

Finally, using $(A\otimes B)(C\otimes D)=AC\otimes BC$ for comformable matrices, we have $\left(\Omega\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)(\Omega^{-1}\otimes \boldsymbol{X}')$ $\left(\Omega\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)=I_n\otimes(\boldsymbol{X}'\boldsymbol{X}).^{-1}\boldsymbol{X}'$. Hence, the MLE estimator is

$$\beta^{\text{mle}}=\left(I_n\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)y \tag{1.27}$$

Note that the MLE estimator coincides with the OLS estimator of equation (1.18), which minimizes the quadratic equation

$$\begin{aligned}
Q^{\text{ols}}(\beta) &= (y-(I_n\otimes \boldsymbol{X})\beta)'(y-(I_n\otimes \boldsymbol{X})\beta)\\
&= y'y-2\beta'(I_n\otimes \boldsymbol{X})'y+\beta'(I_n\otimes \boldsymbol{X})'(I_n\otimes \boldsymbol{X})\beta\\
&= y'y-2\beta'(I_n\otimes \boldsymbol{X}')y+\beta'(I_n\otimes(\boldsymbol{X}'\boldsymbol{X}))\beta.
\end{aligned}$$

The first order condition is

$$-2(I_n\otimes \boldsymbol{X}')y+2(I_n\otimes(\boldsymbol{X}'\boldsymbol{X}))\beta^{\text{ols}}=0\Rightarrow\beta^{\text{ols}}=(I_n\otimes(\boldsymbol{X}'\boldsymbol{X}))^{-1}(I_n\otimes \boldsymbol{X}')y$$

or

$$\beta^{\text{ols}} = \left(I_n \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) y. \tag{1.28}$$

Exercise: the GLS estimator minimizes the quadratic expression

$$Q^{\text{gls}}(\beta) = (y - (I_n \otimes \boldsymbol{X})\beta)' (\Omega \otimes I_T)^{-1} (y - (I_n \otimes \boldsymbol{X})\beta).$$

Prove that $\beta^{\text{gls}} = \beta^{\text{mle}} = \beta^{\text{ols}}$.

Yet another way of writing the likelihood

We now prove yet another way of writing the likelihood function that will be useful below. First note that

$$\Omega^{-1} \otimes I_T = \left(\Omega^{-1/2} \otimes I_T\right)\left(\Omega^{-1/2} \otimes I_T\right)$$

and that

$$\left(\Omega^{-1/2} \otimes I_T\right)' = \Omega^{-1/2} \otimes I_T.$$

We now do some algebra with the quadratic term in (1.20)

$$\begin{aligned}
& (y - (I_n \otimes \boldsymbol{X})\beta)' \left(\Omega^{-1} \otimes I_T\right)(y - (I_n \otimes \boldsymbol{X})\beta) \\
= \; & (y - (I_n \otimes \boldsymbol{X})\beta)' \left(\Omega^{-1/2} \otimes I_T\right)\left(\Omega^{-1/2} \otimes I_T\right)(y - (I_n \otimes \boldsymbol{X})\beta) \\
= \; & \left[\left(\Omega^{-1/2} \otimes I_T\right)'(y - (I_n \otimes \boldsymbol{X})\beta)\right]' \left(\Omega^{-1/2} \otimes I_T\right)(y - (I_n \otimes \boldsymbol{X})\beta) \\
= \; & \left[\left(\Omega^{-1/2} \otimes I_T\right)(y - (I_n \otimes \boldsymbol{X})\beta)\right]' \left(\Omega^{-1/2} \otimes I_T\right)(y - (I_n \otimes \boldsymbol{X})\beta) \\
= \; & \left[\left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes I_T\right)(I_n \otimes \boldsymbol{X})\beta\right]' \left(\left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes I_T\right)(I_n \otimes \boldsymbol{X})\beta\right) \\
= \; & \left[\left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta\right]' \left[\left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta\right]
\end{aligned}$$

where in the third line we used $(AB)' = B'A'$ for $A = (y - (I_n \otimes \boldsymbol{X})\beta)$ and $B = \left(\Omega^{-1/2} \otimes I_T\right)$ and the other properties discussed above.

Now rewrite the term

$$\begin{aligned}
\left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta &= \left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta + \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta^{\text{ols}} - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta^{\text{ols}} \\
&= \left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta^{\text{ols}} + \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\left(\beta^{\text{ols}} - \beta\right) \\
&= \underbrace{\left(\Omega^{-1/2} \otimes I_T\right)y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\beta^{\text{ols}}}_{=W} + \underbrace{\left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)\left(\beta^{\text{ols}} - \beta\right)}_{=Z} \\
&= W + Z
\end{aligned}$$

where β^{ols} is the OLS (and MLE) estimator given in (1.28). Hence, we have

$$\begin{aligned}
(y - (I_n \otimes \boldsymbol{X})\beta)' \left(\Omega^{-1} \otimes I_T\right)(y - (I_n \otimes \boldsymbol{X})\beta) &= (W + Z)'(W + Z) \\
&= W'W + 2Z'W + Z'Z.
\end{aligned}$$

We now consider each of the three terms separately.

Consider first $Z'Z$:

$$\begin{aligned} Z'Z &= \left(\beta^{\text{ols}} - \beta\right)' \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)' \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \left(\beta^{\text{ols}} - \beta\right) \\ &= \left(\beta - \beta^{\text{ols}}\right)' \left(\Omega^{-1/2} \otimes \boldsymbol{X}'\right) \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \left(\beta - \beta^{\text{ols}}\right) \\ &= \left(\beta - \beta^{\text{ols}}\right)' \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \left(\beta - \beta^{\text{ols}}\right). \end{aligned} \tag{1.29}$$

Consider next $W'W$:

$$\begin{aligned} W'W &= \left[\left(\Omega^{-1/2} \otimes I_T\right) y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \beta^{\text{ols}}\right]' \left[\left(\Omega^{-1/2} \otimes I_T\right) y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \beta^{\text{ols}}\right] \\ &= \left[\left(\Omega^{-1/2} \otimes I_T\right) \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right)\right]' \left[\left(\Omega^{-1/2} \otimes I_T\right) \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right)\right] \\ &= \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right)' \left(\Omega^{-1/2} \otimes I_T\right) \left(\Omega^{-1/2} \otimes I_T\right) \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right) \\ &= \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right)' \left(\Omega^{-1} \otimes I_T\right) \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right). \end{aligned}$$

Note that this expression is identical to that leading to formula (1.24) but replacing β by β^{ols}, so the same formula applies and we have

$$\begin{aligned} W'W &= \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right)' \left(\Omega^{-1} \otimes I_T\right) \left(y - (I_n \otimes \boldsymbol{X}) \beta^{\text{ols}}\right) \\ &\quad \text{tr}\left(\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}\right)' \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}\right) \Omega^{-1}\right) \\ &= \text{tr}\left(\boldsymbol{S}^{\text{ols}} \Omega^{-1}\right). \end{aligned} \tag{1.30}$$

where

$$\boldsymbol{S}^{\text{ols}} \equiv \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}\right)' \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}\right) \tag{1.31}$$

is the sum of squared residuals from the OLS estimation of equation (1.16).

Finally, we will show that $Z'W = 0$:

$$\begin{aligned} Z'W &= \left(\beta^{\text{ols}} - \beta\right)' \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right)' \left[\left(\Omega^{-1/2} \otimes I_T\right) y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \beta^{\text{ols}}\right] \\ &= \left(\beta^{\text{ols}} - \beta\right)' \left(\Omega^{-1/2} \otimes \boldsymbol{X}'\right) \left[\left(\Omega^{-1/2} \otimes I_T\right) y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \beta^{\text{ols}}\right] \\ &= \left(\beta^{\text{ols}} - \beta\right)' \left[\left(\Omega^{-1/2} \otimes \boldsymbol{X}'\right) \left(\Omega^{-1/2} \otimes I_T\right) y - \left(\Omega^{-1/2} \otimes \boldsymbol{X}'\right) \left(\Omega^{-1/2} \otimes \boldsymbol{X}\right) \beta^{\text{ols}}\right] \\ &= \left(\beta^{\text{ols}} - \beta\right)' \left[\left(\Omega^{-1/2}\Omega^{-1/2} \otimes \boldsymbol{X}'\right) y - \left(\Omega^{-1/2}\Omega^{-1/2} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{ols}}\right] \\ &= \left(\beta^{\text{ols}} - \beta\right)' \left[\left(\Omega^{-1} \otimes \boldsymbol{X}'\right) y - \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{ols}}\right] \\ &= \left(\beta^{\text{ols}} - \beta\right)' \left[\left(\Omega^{-1} \otimes \boldsymbol{X}'\right) y - \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{mle}}\right], \end{aligned}$$

where the last equality uses that $\beta^{\text{mle}} = \beta^{\text{ols}}$. The MLE first order condition (1.26) then implies that the term in square brackets is zero, implying that

$$Z'W = 0. \tag{1.32}$$

Putting together these three results we obtain

$$\begin{aligned}
& \left(y-\left(I_n\otimes \boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_T\right)\left(y-\left(I_n\otimes \boldsymbol{X}\right)\beta\right) \\
= \;& Z'Z+W'W \\
= \;& \left(\beta-\beta^{\text{ols}}\right)'\left(\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)\left(\beta-\beta^{\text{ols}}\right)+\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right). \qquad (1.33)
\end{aligned}$$

Therefore, we can write the likelihood function (1.20) as

$$L\left(\beta,\Omega|Y\right)=\frac{\left|\Omega\right|^{-T/2}}{\left(2\pi\right)^{nT/2}}\exp\left[-\frac{1}{2}\left(\beta-\beta^{\text{ols}}\right)'\left(\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)\left(\beta-\beta^{\text{ols}}\right)\right]\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]. \qquad (1.34)$$

This is a standard decomposition of the likelihood function. The first exponent is the kernel of a multivariate normal with mean β^{ols} and covariance $\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. The second exponent looks something like a Wishart distribution. Recall that $\boldsymbol{X}$ is $T\times k$ with $k=np+1$, $\boldsymbol{Y}$ is $T\times n$, so that $\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}$ is $T\times n$ so that $\boldsymbol{S}^{\text{ols}}=\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}\right)'\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{\text{ols}}\right)$ is $n\times n$.

The term $\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]$ looks like the kernel of the inverted Wishart distribution (1.5) with scale matrix $\Psi=\boldsymbol{S}^{\text{ols}}$ and random matrix $Z=\Omega$. So we need to keep track and fix some of the terms in (1.34). As mentioned before, the first exponent is the kernel of a multivariate normal with mean β^{ols} and covariance $\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$, so in the likelihood we must have the term $\left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-1/2}$ but in (1.34) we have $\left|\Omega\right|^{-T/2}$. Also, recall that $\boldsymbol{X}'\boldsymbol{X}$ (and $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$) is a $k\times k$ matrix. Then, using the property of the determinant of a Kronecker product we have

$$\left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|=\left|\Omega\right|^{k}\left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{n}$$

So let's write

$$\begin{aligned}
\left|\Omega\right|^{-\frac{T}{2}} &= \left|\Omega\right|^{-\frac{k}{2}}\left|\Omega\right|^{\frac{k-T}{2}} \\
&= \left|\Omega\right|^{-\frac{k}{2}}\left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-\frac{n}{2}}\left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{\frac{n}{2}}\left|\Omega\right|^{-\frac{T-k}{2}} \\
&= \left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{\frac{n}{2}}\left|\Omega\right|^{-\frac{k}{2}}\left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-\frac{n}{2}}\left|\Omega\right|^{-\frac{T-k}{2}} \\
&= \left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{\frac{n}{2}}\left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-1/2}\left|\Omega\right|^{-\frac{T-k}{2}}.
\end{aligned}$$

Also, since β is a $kn\times 1$ vector, in the multivariate normal we must have the term $\left(2\pi\right)^{kn/2}$ but we have $\left(2\pi\right)^{Tn/2}$. But this is easy

$$\left(2\pi\right)Tn/2=\left(2\pi\right)^{kn/2}\left(2\pi\right)^{(T-k)n/2}.$$

With these adjustments we can write (1.34) as

$$\begin{aligned}
L(\beta,\Omega|Y) &= \left|\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{\frac{n}{2}}\frac{\left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-1/2}}{\left(2\pi\right)^{kn/2}}\exp\left[-\frac{1}{2}\left(\beta-\beta^{\text{ols}}\right)'\left(\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)\left(\beta-\beta^{\text{ols}}\right)\right] \\
&\quad\times\frac{\left|\Omega\right|^{-\frac{T-k}{2}}}{2\pi)^{(T-k)n/2}}\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right].
\end{aligned}$$

Clearly,

$$\frac{\left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-1/2}}{\left(2\pi\right)^{kn/2}}\exp\left[-\frac{1}{2}\left(\beta-\beta^{\text{ols}}\right)'\left(\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)\left(\beta-\beta^{\text{ols}}\right)\right]$$

is the density of a multivariate normal distribution for $\beta\in R^{kn}$ with mean β^{ols} and covariance matrix $\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$, so that $\beta\sim N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)$.

As for the second term, since $\boldsymbol{S}^{\text{ols}}$ is $n\times n$, we only need to make a small adjustment and write $T-k=(T-k-n-1)+n+1$

$$|\Omega|^{-\frac{T-k}{2}}\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]=|\Omega|^{-\frac{(T-k-n-1)+n+1}{2}}\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]$$

which is the kernel of an inverted Wishart distribution (1.5) for Ω with scale matrix $\Psi=\boldsymbol{S}^{\text{ols}}$ and $T-k-n-1$ degrees of freedom.

All in all, we proved that the likelihood function is proportional to the product of a normal distribution for β given Ω ,$\beta\sim N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)$, and an inverse Wishart distribution for $\Omega\sim iW\left(\boldsymbol{S}^{\text{ols}},T-k-n-1\right)$, and hence the likelihood function can finally be written as

$$L\left(\beta,\Omega|Y\right)\propto N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)\times iW\left(\boldsymbol{S}^{\text{ols}},T-k-n-1\right). \tag{1.35}$$

Equation (1.35) is very important as it will allows us to appropriately choose conjugate priors to obtain conditional posteriors in closed form. As we will see below, a Normal-Wishart prior conjugates the two blocks of the likelihood function.

Priors for VARs

In this section we will consider useful priors for our VAR(p) (1.14). Namely, we consider

a) A Normal prior for β assuming Ω is known (Theil mixed estimator).

b) A noninformative prior for both β and Ω (Jeffreys prior).

c) A Normal prior for β and a non-informative prior for Ω.

d) A Normal prior for β and an independent Wishart for Ω^{-1}.

Case 1. Normal prior for β for a given Ω

In this case we assume that we know Ω. To implement the procedure, we set Ω to its OLS estimated value. Let's assume the following prior for β

$$\beta\sim N\left(\beta_0,V_0\right),$$

so that

$$p\left(\beta\right)\propto|V_0|^{-0.5}\exp\left(-\frac{1}{2}\left(\beta-\beta_0\right)'V_0^{-1}\left(\beta-\beta_0\right)\right).$$

Using equation (1.20) we have

$$L\left(\beta,\Omega|Y\right)=\frac{\left|\Omega\right|^{-T/2}}{\left(2\pi\right)^{nT/2}}\exp\left[-\frac{1}{2}\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_{T}\right)\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)\right].$$

Hence the posterior satisfies

$$\begin{aligned}p\left(\beta|\Omega,Y\right) &\propto p\left(\beta\right)L\left(\beta,\Omega|Y\right)\\ &\propto \exp\left[-\frac{1}{2}\left(\beta-\beta_{0}\right)'V_{0}^{-1}\left(\beta-\beta_{0}\right)\right]\exp\left[-\frac{1}{2}\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_{T}\right)\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)\right]\\ &\propto \exp\left[-\frac{1}{2}\left[\left(\beta-\beta_{0}\right)'V_{0}^{-1}\left(\beta-\beta_{0}\right)+\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_{T}\right)\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)\right]\right].\end{aligned}$$

The term in the exponent contains a sum of two quadratic expressions in β which will also be another quadratic expression in β. Let's decompose the two terms. First,

$$\left(\beta-\beta_{0}\right)'V_{0}^{-1}\left(\beta-\beta_{0}\right)=\beta'V_{0}^{-1}\beta-2\beta'V_{0}^{-1}\beta_{0}+\beta_{0}'V_{0}^{-1}\beta_{0}.$$

Also, when we derived the MLE estimation of β we proved that

$$\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_{T}\right)\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)=y'\left(\Omega^{-1}\otimes I_{T}\right)y-2y'\left(\Omega^{-1}\otimes\boldsymbol{X}\right)\beta+\beta'\left(\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)\beta$$

Hence,

$$\begin{aligned}&\left(\beta-\beta_{0}\right)'V_{0}^{-1}\left(\beta-\beta_{0}\right)+\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_{T}\right)\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)\\ =&\beta'V_{0}^{-1}\beta-2\beta'V_{0}^{-1}\beta+\beta_{0}'V_{0}^{-1}\beta_{0}+y'(\Omega^{-1}\otimes I_{T})y-2y'(\Omega^{-1}\otimes\boldsymbol{X})+\beta'(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X}))\beta\\ =&\beta'\left[V_{0}^{-1}+\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right]\beta-2\left[\beta_{0}'V_{0}^{-1}+y'(\Omega^{-1}\otimes\boldsymbol{X})\right]\beta+\beta_{0}'V_{0}^{-1}\beta_{0}+y'(\Omega^{-1}\otimes I_{T})y\end{aligned}\tag{1.36}$$

The last two terms of the sum are constants.

We now write these two expressions as a quadratic equation in β of the form

$$\left(\beta-\beta_{1}\right)V_{1}^{-1}\left(\beta-\beta_{1}\right)$$

where C is a constant. Note that

$$\left(\beta-\beta_{1}\right)V_{1}^{-1}\left(\beta-\beta_{1}\right)=\beta'V_{1}^{-1}\beta-2\beta_{1}'V_{1}^{-1}\beta+\beta_{1}'V_{1}^{-1}\beta_{1}.\tag{1.37}$$

Comparing (1.36) with (1.37) we can match coefficients (this is called completing the square). First, define

$$V_{1}^{-1}=V_{0}^{-1}+\left(\Omega^{-1}\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)\right)$$

and

$$\beta_{1}'V_{1}^{-1}=\beta_{0}'V_{0}^{-1}+y'\left(\Omega^{-1}\otimes\boldsymbol{X}\right)\Rightarrow\beta_{1}'=\left[\beta_{0}'V_{0}^{-1}+y'\left(\Omega^{-1}\otimes\boldsymbol{X}\right)\right]V_{1}$$

or

$$\beta_{1}=V_{1}\left[V_{0}^{-1}\beta_{0}+\left(\Omega^{-1}\otimes\boldsymbol{X}'\right)y\right]$$

This implies that

$$\left(\beta-\beta_{0}\right)'V_{0}^{-1}\left(\beta-\beta_{0}\right)+\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)'\left(\Omega^{-1}\otimes I_{T}\right)\left(y-\left(I_{n}\otimes\boldsymbol{X}\right)\beta\right)=\left(\beta-\beta_{1}\right)V_{1}^{-1}\left(\beta-\beta_{1}\right)+C$$

where

$$V_1 = \left[V_0^{-1} + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\right]^{-1} \tag{1.38}$$

$$\beta_1 = V_1 \left[V_0^{-1}\beta_0 + \left(\Omega^{-1} \otimes \boldsymbol{X}'\right) y\right] \tag{1.39}$$

and C is a constant (i.e. does not depend on β).

With all this algebra we conclude that the posterior density satisfies

$$p(\beta|\Omega, Y) \propto N(\beta_1, V_1). \tag{1.40}$$

But we can do a little more. Take the term $(\Omega^{-1} \otimes \boldsymbol{X}')\, y$ in (1.39) and recall that the OLS estimate satisfies

$$\left(\Omega^{-1} \otimes \boldsymbol{X}'\right) y = \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{ols}}.$$

Hence we can write (1.39) as

$$\begin{aligned}
\beta_1 &= V_1 \left[V_0^{-1}\beta_0 + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{ols}}\right] \\
&= \left[V_0^{-1} + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\right]^{-1} \left[V_0^{-1}\beta_0 + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{ols}}\right] \\
&= \left[V_0^{-1} + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\right]^{-1} V_0^{-1}\beta_0 + \left[V_0^{-1} + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\right]^{-1} \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right) \beta^{\text{ols}} \\
&= W\beta_0 + (I - W)\, \beta^{\text{ols}},
\end{aligned}$$

where the weighting matrix $W = \left[V_0^{-1} + (\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}))\right]^{-1} V_0^{-1}$. This implies that the posterior is a weighted average of the prior β_0 and the OLS estimate β^{ols}.

To implement this procedure, we proceed as follows:

`i) Choose the parameters of the prior distribution of` β : β_0 `and` V_0.

`ii) Set` $\Omega = \Omega^{\text{ols}}$, `the OLS estimate of the covariance matrix.`

`iii) Draw from a Normal distribution with mean` β_1 `and covariance matrix`

$$\left[V_0^{-1} + \Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right]^{-1}.$$

Although simple, this procedure still requires us to choose a large number of parameters for β_0 (nk parameters) and for its covariance matrix, V_0 ($kn(kn-1)/2$ parameters). This is too much. The literature has considered several approaches to simplify these choices. One prior that is quite popular in applied work is the *Minnesota* or *Litterman* prior, which we discuss next.

The Minnesota Prior

The Minnesota, or Litterman, prior, is a special case of Case 1 in which the prior parameters β_0 and V_0 are functions of a small number of hyperparameters. The original proposal is typically used for variables in levels and shrinks the VAR estimates toward a random walk for each variable. When the variables are in growth rates, it is common to specify the prior of the parameters in β to have all mean zero. The Minnesota prior has been useful in forecasting persistent economic time series. The traditional implementation for persistence series is as follows:

a) For each equation of the VAR, set the prior mean of the first lag of the dependent variable to one and set the prior mean of all other slope coefficients to zero. That is, the prior is that each variable follows an independent random walk,

$$y_{i,t} = y_{i,t-1} + v_{it}.$$

b) Set the prior variance of the ij^{th} element of the matrix D_ℓ, denoted by $v_{ij,\ell}^D$ to

$$v_{ij,\ell}^D = \begin{cases} \frac{\lambda_1}{\ell^{\lambda_3}} & \text{if } i = j \\ \frac{\lambda_1 \lambda_2}{\ell^{\lambda_3}} \frac{\sigma_i^2}{\sigma_j^2} & \text{if } i \neq j \end{cases}.$$

c) Set the prior variances of the intercept c (or other exogenous variables) as

$$v_i^c = \lambda_1 \lambda_4.$$

where λ_4 is usually a large number.

That is, the assumption is that the prior covariance matrix V_0 is a diagonal matrix that depends only on 4 parameters: λ_1, λ_2, λ_3, and λ_4. In this matrix, the variance of the coefficient in the i-th variable's own lags are set to $\lambda_1/\ell^{\lambda_3}$. That is, λ_1 is the variance of the coefficient on the first own lag and λ_3 is a decay parameter that controls how fast the variance on higher own lags shrinks toward zero. For variables j other than the i-th, the parameter λ_2 controls the relative tightness in the other lags compared to the own lag, so that a smaller λ_2 increases the relative tightness of the other lags. The term σ_j^2/σ_i^2 is a scale factor to account for the different variances of the variables. σ_i^2 is the i-th diagonal element of Ω. Finally, λ_4 measures the tightness of the constant or other exogenous variables in the VAR. Since normally we do not want to impose tight priors on these parameters, we usually set λ_4 to a large number (infinity).

Example: Consider a VAR with two endogenous variables and two lags along with a constant,

$$y_t = c + D_1 y_{t-1} + D_2 y_{t-2} + v_t.$$

In this case, each equation involves $k = np + 1 = 2 \times 2 + 1 = 5$ coefficients to estimate, for a total of $kn = 10$ parameters in β. Recall that

$$\beta = \text{vec}\left(\begin{bmatrix} D_1' \\ D_2' \\ c' \end{bmatrix}\right) = \text{vec}\left(\begin{bmatrix} D_1(1,1) & D_1(2,1) \\ D_1(1,2) & D_1(2,2) \\ D_2(1,1) & D_2(2,1) \\ D_2(1,2) & D_2(2,2) \\ c_1 & c_2 \end{bmatrix}\right) = \begin{bmatrix} D_1(1,1) \\ D_1(1,2) \\ D_2(1,1) \\ D_2(1,2) \\ c_1 \\ D_1(2,1) \\ D_1(2,2) \\ D_2(2,1) \\ D_2(2,2) \\ c_2 \end{bmatrix}$$

We set the Minnesota prior for the mean as

$$\beta_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} D_1(1,1) \\ D_1(1,2) \\ D_2(1,1) \\ D_2(1,2) \\ c_1 \\ D_1(2,1) \\ D_1(2,2) \\ D_2(2,1) \\ D_2(2,2) \\ c_2 \end{bmatrix}$$

The prior variance of β is set to

$$V_0 = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1\lambda_2\left(\frac{\sigma_1}{\sigma_2}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\lambda_1}{2^{\lambda_3}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\lambda_1\lambda_2}{2^{\lambda_3}}\left(\frac{\sigma_1}{\sigma_2}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_1\lambda_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_1\lambda_2\left(\frac{\sigma_2}{\sigma_1}\right)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\lambda_1\lambda_2}{2^{\lambda_3}}\left(\frac{\sigma_2}{\sigma_1}\right)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\lambda_1}{2^{\lambda_3}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1\lambda_4 \end{bmatrix}.$$

Case 2: Noninformative prior for β and Ω

Here we consider a "diffuse" prior for β. Jeffreys (1961) proposed a rule for generating priors that are noninformative about the parameter of interest and retain certain useful properties. The Jeffreys' (or diffuse) prior is proportional to the square root of the determinant of the Fisher information matrix. In the case of a VAR with n variables, the diffuse prior is

$$p(\beta, \Omega) = p(\beta)\, p(\Omega)$$

with

$$p(\beta) = \text{constant} \tag{1.41}$$

$$p(\beta, \Omega) \propto |\Omega|^{-(n+1)/2}. \tag{1.42}$$

Then, the posterior distribution is

$$p(\beta, \Omega|Y) = L(\beta, \Omega|Y)\, p(\beta)\, p(\Omega)$$

Write the likelihood as in (1.34)

$$L(\beta, \Omega|Y) \propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\left(\beta - \beta^{\text{ols}}\right)'\left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\left(\beta - \beta^{\text{ols}}\right)\right] \times \exp\left[-\frac{1}{2} tr\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right].$$

Above we showed that the likelihood can be written as

$$L(\beta,\Omega|Y) \propto \left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\beta-\beta^{\text{ols}}\right)'\left(\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)^{-1}\left(\beta-\beta^{\text{ols}}\right)\right]$$
$$\times|\Omega|^{-\frac{(T-k-n-1)+n+1}{2}}\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right].$$

Then the posterior is

$$\begin{aligned}p(\beta,\Omega|Y) \propto& \left|\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\beta-\beta^{\text{ols}}\right)'\left(\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)^{-1}\left(\beta-\beta^{\text{ols}}\right)\right]\\ &\times|\Omega|^{-\frac{(T-k-n-1)+n+1}{2}}|\Omega|^{-(n+1)/2}\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]\\ \propto& N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)|\Omega|^{-\frac{(T-k)+n+1}{2}}\exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]\\ \propto& N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)\times iW\left(\Omega|\boldsymbol{S}^{\text{ols}},T-k\right).\end{aligned} \tag{1.43}$$

Because the normal density integrates to 1, integrating over β we have

$$p\left(\Omega|Y\right)=\int_{\beta}p\left(\beta,\Omega|Y\right)d\beta=iW\left(\Omega|\boldsymbol{S}^{\text{ols}},T-k\right).$$

Using the definition of conditional probability

$$p\left(\beta,\Omega|Y\right)=p\left(\beta|\Omega,Y\right)P\left(\Omega|Y\right)$$

which implies, together with (1.43), that

$$p\left(\beta|\Omega,Y\right)=\frac{p\left(\beta,\Omega|Y\right)}{P\left(\Omega|Y\right)}\propto N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right).$$

Summarizing results, we have that

$$p\left(\beta|\Omega,Y\right)=N\left(\beta^{\text{ols}},\Omega\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right). \tag{1.44}$$

$$p\left(\Omega|Y\right)=iW\left(\Omega,\boldsymbol{S}^{\text{ols}},T-k\right) \tag{1.45}$$

It is possible to compute $p(\beta|Y)$ in closed form as done in Zellner (1971). But we can also simulate the posterior $p(\beta|Y)$ using a version of the Gibbs Sampler.

Algorithm (Gibbs Sampler): Choose a large N and set $j=1$. Then iterate on the following loop:

a) Draw Ω^j from $iW\left(\Omega|\boldsymbol{S}^{\text{ols}},T-k\right)$.

b) Draw β^j from $N(\beta^{\text{ols}}|\Omega^j\otimes\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1})$.

c) Store $\{\beta^j,\Omega^j\}$, set $j=j+1$, and return to step a) while $j<N$.

Case 3: Normal prior for β and non-informative prior for Ω

Here we consider the following prior

$$p(\beta, \Omega) = p(\beta)\, p(\Omega)$$

with

$$p(\beta) \sim N(\beta_0, V_0) \tag{1.46}$$

$$p(\Omega) \propto |\Omega|^{-(n+1)/2}. \tag{1.47}$$

As in the previous case, we will draw from the posterior using the Gibbs Sampler.

To that end, we first compute $p(\Omega|\beta, Y)$ assuming that β is known (i.e. conditioning on β). Using the likelihood representation (1.25)

$$\begin{aligned} L(\beta, \Omega|Y) =& \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V}\right)\right] \\ \propto& |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left((\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})\Omega^{-1}\right)\right] \end{aligned}$$

The joint posterior is then

$$\begin{aligned} p(\Omega|\beta, Y) &\propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left((\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})\Omega^{-1}\right)\right] |\Omega|^{-(n+1)/2} \\ &\propto |\Omega|^{-(T+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}\Omega^{-1}\right)\right], \end{aligned}$$

where

$$\boldsymbol{S} = (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}). \tag{1.48}$$

Therefore, the conditional distribution

$$p(\Omega|\beta, Y) = iW(\boldsymbol{S}, T)$$

is an inverse Wishart with scale matrix $\boldsymbol{S}$ and T degrees of freedom.

Now suppose that Ω is known and, following the steps derived for Case 1, we conclude that

$$p(\beta|\Omega, Y) = N(\beta_1, V_1).$$

where

$$\begin{aligned} V_1 =& \left[V_0^{-1} + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\right]^{-1} \\ \beta_1 =& V_1\left[V_0^{-1}\beta_0 + \left(\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right)\beta^{\mathrm{ols}}\right]. \end{aligned}$$

Summarizing, we have the conditional posteriors $p(\Omega|\beta, Y)$ and $p(\beta|\Omega, Y)$, so we can use the Gibbs Sampler.

Algorithm (Gibbs Sampler): `Choose` N `large, an arbitrary` β^0`, and set` $j = 1$`. Then iterate on the following loop:`

a) Draw Ω^j from the conditional posterior $iW(\boldsymbol{S}^{j-1}, T)$, where

$$\boldsymbol{S}^{j-1} = \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{j-1}\right)' \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{j-1}\right)$$

is the sum of squared residuals using β^{j-1} as the VAR slope parameters.

b) Given Ω^j obtained in the previous step, draw $\beta^j \sim N(\beta_1, V_1)$ using the above formulas replacing Ω by Ω^j.

c) Set $j \to j+1$ and return to step a). Repeat many times and discard an initial burn in sample.

Case 4: The independent Normal-Wishart prior

This case is a Normal prior for β and inverse Wishart prior for Ω. The steps are quite similar to Case 3 with some minor modifications. Here we consider

$$p(\beta, \Omega) = p(\beta)p(\Omega)$$

with

$$p\left(\beta\right) \sim N\left(\beta_0, V_0\right) \propto \exp\left[-\frac{1}{2}\left(\beta - \beta^*\right)' V_\beta^{-1}\left(\beta - \beta^*\right)\right]$$

and

$$\Omega \sim iW\left(\boldsymbol{S}_0, v_0\right) \propto \left|\Omega\right|^{-\frac{v_0+n+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}_0\Omega^{-1}\right)\right].$$

As above, first compute $p\left(\Omega|\beta, Y\right)$ assuming that β is known by using the likelihood representation (1.25)

$$L\left(\beta, \Omega|Y\right) \propto \left|\Omega\right|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}\Omega^{-1}\right)\right]$$

where $\boldsymbol{S}$ is given by (1.48). Then the conditional posterior is

$$\begin{aligned} p(\Omega|\beta, Y) \propto & \left|\Omega\right|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}\Omega^{-1}\right)\right] \left|\Omega\right|^{-\frac{v^*+n+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}^*\Omega^{-1}\right)\right] \\ \propto & \left|\Omega\right|^{-(T+v_0+n+1)/2} \exp\left[-\frac{1}{2}\left[\mathrm{tr}\left[\left(\boldsymbol{S} + \boldsymbol{S}_0\right)\Omega^{-1}\right]\right]\right] \\ \propto & iW\left(\boldsymbol{S} + \boldsymbol{S}_0, T + v_0\right). \end{aligned}$$

This is, of course, the kernel of an inverted Wishart distribution with $T + v_0$ degrees of freedom and $\boldsymbol{S} + \boldsymbol{S}_0$ scale coefficient matrix.

Now, conditional on Ω, the posterior of β we already know:

$$p\left(\beta|\Omega, Y\right) = N\left(\beta_1, V_1\right)$$

where

$$\begin{aligned} V_1 = & \left[V_0^{-1} + \left(\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right)\right]^{-1} \\ \beta_1 = & V_1\left[V_0^{-1}\beta_0 + \left(\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right)\beta^{\mathrm{ols}}\right]. \end{aligned}$$

Summarizing, we have the conditional posteriors $p(\Omega|\beta, Y)$ and $p(\beta|\Omega, Y)$, so we can use the Gibbs Sampler. The algorithm is copied almost verbatim from that in the previous section, with the only difference in step 2

Algorithm: Choose a large N, an arbitrary β^0, and set $j = 1$. Then iterate on the following loop:

a) Draw Ω^j from the conditional posterior $iW(\boldsymbol{S}^{j-1} + \boldsymbol{S_0}, T + v_0)$, where

$$\boldsymbol{S}^{j-1} = \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{\boldsymbol{j-1}}\right)' \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{\boldsymbol{j-1}}\right)$$

is the sum of squared residuals using β^{j-1} at the VAR slope parameters.

b) Given Ω^j obtained in the previous step, draw $\beta^j \sim N(\beta_1, V_1)$ using the above formulas replacing Ω by Ω^j.

c) Set $j \to j+1$ and return to step a). Repeat many times and discard an initial burn in sample.

Numerical approximation of DSGE models

Constantino Hevia
Universidad Torcuato Di Tella

This note discusses the numerical approximation of DSGE (dynamic, stochastic, general equilibrium) models. The equilibrium conditions of a model can be cast in the form of a system of non-linear stochastic difference equations. The solution to this system determines the solution of the model. In this note we discuss how to approximate the nonlinear system using a linear approximation around the non-stochastic steady state. To build intution, we start with the example of a simple RBC model. The method, however, is general and can be applied to a wide range of models.

A REAL BUSINESS CYCLE (RBC) MODEL

This is a model of a closed economy with a large number of identical households with size normalized to one. We focus on the behavior of the "representative" household. There are no market failures. The First Welfare Theorem holds and the competivite equilibrium is Pareto efficient. We thus use the Second Welfare Theorem to find the equilibrium allocation by solving the planner's problem of maximizing the utility of the representative household subject to feasibility constraints. Prices, if needed, can be computed from the appropriate conditions of the associated competitive equilibrium.

A representative household has preferences over consumption $\{c_t\}$ and leisure $\{h_t\}$ represented by the utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t U(c_t, h_t) \tag{1}$$

where $U(c,h)$ is increasing in each argument and concave, $\beta \in (0,1)$ is a discount factor, and E_0 is the expectation operator conditional on information at time $t=0$.

The feasibility constraint of the economy is

$$c_t + i_t = y_t,$$

where i_t is investment and y_t is output. Output is produced with capital, k_t, and labor, l_t, according to the production function

$$y_t = A_t k_t^{\alpha} l_t^{1-\alpha},$$

where A_t is the level of technology. Technology evolves according to

$$\log(A_{t+1}) = \rho \log(A_t) + \varepsilon_{t+1},$$

where $|\rho| < 1$ and $\varepsilon_t \sim N\left(0, \sigma_{\varepsilon}^2\right)$.

The stock of capital evolves according to

$$k_{t+1} = (1-\delta)\, k_t + i_t.$$

With this notation, k_t is the stock of capital installed at the beginning of period t that was chosen in period $t-1$. At time t, agents choose k_{t+1}.

If we normalize the total amount of time for consuming leisure or working to one, labor market feasibility requires

$$l_t + h_t = 1.$$

The initial stock of capital k_0 and level of technology A_0 are given.

Using these constraints, the planner's problem can be written as

$$\max E_0 \sum_{t=0}^{\infty} \beta^t U\left(c_t, 1-l_t\right)$$

subject to

$$\begin{aligned} c_t + k_{t+1} &= A_t k_t^{\alpha} l_t^{1-\alpha} + (1-\delta)\, k_t && (2)\\ \log A_{t+1} &= \rho \log A_t + \varepsilon_{t+1} && (3)\\ & \quad k_0 \text{ and } A_0 \text{ given.} \end{aligned}$$

The Lagrangian of this problem is

$$\max_{\{c_t, l_t, k_{t+1}, \lambda_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \left[U\left(c_t, 1-l_t\right) - \lambda_t \left(c_t + k_{t+1} - A_t k_t^{\alpha} l_t^{1-\alpha} - (1-\delta)\, k_t\right)\right].$$

where $\beta^t \lambda_t$ are the (stochastic) Lagrange multipliers, and k_0 and A_0 are given.

The first order conditions with respect to c_t, l_t, k_{t+1}, and λ_t are given by

$$U_c\left(c_t, 1-l_t\right) - \lambda_t = 0 \tag{4}$$

$$-U_h\left(c_t, 1-l_t\right) - \lambda_t (1-\alpha) A_t k_t^{\alpha} l_t^{-\alpha} = 0 \tag{5}$$

$$\beta E_t \left[\lambda_{t+1} \left(\alpha A_{t+1} k_{t+1}^{\alpha-1} l_{t+1}^{1-\alpha} + 1 - \delta\right)\right] - \lambda_t = 0 \tag{6}$$

$$c_t + k_{t+1} - A_t k_t^{\alpha} l_t^{1-\alpha} - (1-\delta)\, k_t = 0, \tag{7}$$

where U_c and U_h denote the partial derivative of U with respect to consumption and leisure, respectively.

There is another necessary condition of the planner that is called the "transversality" condition, which

is given by

$$\lim_{T\to\infty} E_0\left[\beta^T \lambda_T k_{T+1}\right] = 0.$$

This condition imposes a constraint on the rate of growth of the stock of capital and is analogous to the requirement that households do not end their lives with a positive stock of capital in a finite horizon version of this model.

We look for solutions in the form of time-invariant "policy functions." To do this, we need to categorize all variables either as state or control variables. The set of state variables (also called predetermined variables) is the minimum set of variables that completely characterizes the state of the economy at the beginning of time t. These are the variables that are taken as given at the beginning of time t and that suffice to determine the current and future evolution of the economy. In some cases it is not obvious how to identify the state variables. Here it is simple: the state variables are the stock of capital k_t and the level of technology A_t; we denote by $x_t = (k_t, A_t)'$ the vector of state variables. The "jump," "control," or "non-predetermined" variables are all those variables that are chosen by the planner and are not predetermined at time t. In this example, the jump variables are c_t, l_t, and λ_t. We denote the set of control variables as $y_t = (c_t, l_t, \lambda_t)'$. It is possible to simplify the system and, for example, get rid of the multiplier λ_t from the above first order conditions. How we choose to solve the model is a matter of convenience. In this note, however, we will keep all the first order conditions derived above because it will be somewhat easier when we perform the linearization of the equilibrium conditions.

Thus, with some abuse of notation, we look for policy functions of the form

$$\begin{aligned} c_t &= \mathsf{c}(k_t, A_t) \\ l_t &= \mathsf{l}(k_t, A_t) \\ \lambda_t &= \lambda(k_t, A_t) \\ k_{t+1} &= \mathsf{k}(k_t, A_t). \end{aligned}$$

where $\mathsf{c}(k_t, A_t)$, $\mathsf{l}(k_t, A_t)$, $\lambda(k_t, A_t)$, and $\mathsf{k}(k_t, A_t)$ are time invariant functions of the state.

Sometimes it is useful to separate the set of state variables into exogenous and endogenous state variable. The former evolve exogenously and cannot be changed by anyone. In contrast, the latter can be influenced by the agents' choices. In the RBC example A_t is the exogenous state variable and k_t is the endogenous state variable.

Note that we can write the equilibrium conditions of the model as a system of equations of the form

$$E_t\left[f\left(x_{t+1}, y_{t+1}, x_t, y_t\right)\right] = \bar{0}, \tag{8}$$

where $\bar{0}$ is a vector of zeros of dimension 5 (one for each equation) and $f : R^{2\times2+2\times3} \to R^5$ is given by

$$f\left(x_{t+1}, y_{t+1}, x_t, y_t\right) = \begin{bmatrix} U_c\left(c_t, 1-l_t\right) - \lambda_t \\ -U_h\left(c_t, 1-l_t\right) - \lambda_t\left(1-\alpha\right) A_t k_t^{\alpha} l_t^{-\alpha} \\ \beta\lambda_{t+1}\left(\alpha A_{t+1} k_{t+1}^{\alpha-1} l_{t+1}^{1-\alpha} + 1 - \delta\right) - \lambda_t \\ c_t + k_{t+1} - A_t k_t^{\alpha} l_t^{1-\alpha} - \left(1-\delta\right) k_t \\ \log\left(A_{t+1}\right) - \rho\log\left(A_t\right) - \varepsilon_{t+1} \end{bmatrix}$$

In general, the equilibrium conditions of a wide range of models can be written as (8), which shows that the equilibrium allocation (and prices if we were solving the competitive equilibrium directly without using the Second Welfare Theorem) solves a nonlinear system of stochastic difference equations.

IMPORTANT RESULTS FROM LINEAR ALGEBRA

Here we compile some useful results from linear algebra. Let A be a matrix. We use a_{ij} to denote element (i, j) of A. If A is a square matrix of complex numbers, we denote by A^H the Hermitian transpose of A. The Hermitian transpose is the generalization of the transpose of a real matrix and means that we first transpose A and next take the complex conjugate of the resulting numbers.

Definition 1: A square matrix A of complex number is said to be unitary if $A^H A = AA^H = I$, where I is the identity matrix.

Comment: Clearly, the inverse of a unitary matrix A exists and equals A^H.

Result 1: A square matrix A is invertible if and only if all its eigenvalues are different from zero.

Definition 2: A square matriz A is upper triangular if its entries below the main diagonal are zero.

Result 2: If A is upper triangular with non-zero entries on its main diagonal, then A is invertible.

The proof follows from the observation that the eigenvalues of a triangular matrix are its diagonal entries and by Result 1.

Result 3: If A is upper triangular and invertible, then A^{-1} is upper triangular. Moreover, the diagonal elements of A^{-1} are the reciprocal of the diagonal elements of A, that is, element (i, i) of A^{-1} is $1/a_{ii}$.

Result 4: If A and B are $n \times n$ upper triangular matrices, then AB is also upper triangular.

The following is a key theorem that we use in the rest of this note.

Theorem 1: *(QZ decomposition): Let A and B be $n \times n$ matrices. If there is a complex number z such that $\det\left(B - Az\right) \neq 0$, then there are matrices Q, Z, S, and T such that:*

1. *Q and Z are unitary, i.e.* $Q^H Q = QQ^H = I$ and $Z^H Z = ZZ^H = I$

2. *T and S are upper triangular*

3. *The matrices Q, Z, S, and T satisfy*

$$\begin{aligned} QAZ &= S \\ QBZ &= T, \end{aligned}$$

4. *There is no index i such that $s_{ii} = t_{ii} = 0$, and*

5. *The matrices Q, Z, S, and T can be chosen in such a way as to make the diagonal entries s_{ii} and t_{ii} appear in any desired order.*

Remark 1: in Theorem 1, the ratios t_{ii}/s_{ii} are called the *generalized eigenvalues* of the matrix pair (A, B) because they can be shown to solve the following eigenvalue problem: find the λ that satisfies $\det(B - A\lambda) = 0$. By convention, we say that $s_{ii} = 0$ corresponds to an infinite generalized eigenvalue.

Remark 2: Theorem 1 is also called the *Generalized Schur Decomposition.*

SOLUTION TO FIRST ORDER APPROXIMATIONS OF DSGE MODELS

Let $x_t \in R^n$ denote a vector of state variables (predetermined variables) and $y_t \in R^m$ a vector of non-state variables (control variables).[1] The vector of state variables includes endogenous state variables, like the stock of capital, and exogenous state variables, like the evolution of technology shocks.

The equilibrium conditions of a wide variety of dynamic economic models can be expressed as the following system of nonlinear stochastic difference equations

$$E_t \left[f\left(x_{t+1}, y_{t+1}, x_t, y_t \right) \right] = \bar{0}, \tag{9}$$

where E_t denotes the mathematical expectation conditional on the information available at time t, the function $f : R^{2n+2m} \to R^{n+m}$, contains all the equilibrium conditions of the model, like Euler equations, feasibiliy conditions, state evolution equations, and so forth, and $\bar{0}$ denotes a vector of zeros of dimension $n + m$. The function f has image on R^{n+m}. The reason for this is that there is one equation (such as a first order condition or a law of motion) for each (control or state) variable: since there are $n + m$ variables, there must be $n + m$ equations.

Please note that there are other ways of writing the equilibrium conditions of the model, as you will notice if you read different papers in the literature. For example, one can also write equation (9) as a

[1] This part of the note follows Schmitt-Grohé and Uribe (2017), Appendix to chapter 4. See also Klein (2000) and Gomme and Klein (2011).

second order stochastic difference equation in the state variables alone and no control variables; that is, as a nonlinear function of x_t, x_{t+1}, and x_{t+2} only. This means that we need two sets of conditions to pin down the solution to (9). The first set of conditions are the initial conditions of the state variables at time $t = 0$, x_0 (like the initial stock of capital). We need to come up with another condition, for otherwise the second order difference equation does not uniquely pin down the solution. This additional condition is usually a "transversality" condition that requires the state variables to be bounded in an appropriate sense. By focusing on bounded solutions, the method that we develop below automatically imposes the required transversality condition.

We further assume that the state vector x_t can be partitioned as $x_t = \left(x'_{1,t}\ x'_{2,t}\right)'$, where $x_{1,t}$ contains all endogenous state variables and $x_{2,t}$ contains all exogenos state variables.[2] We assume that the dimensions of $x_{1,t}$ and $x_{2,t}$ are, respectively n_1 and n_2, where $n_1 + n_2 = n$. Furthermore, we will focus on cases where the exogenous state variables evolve according to the following stochastic process

$$x_{2,t+1} = \Lambda x_{2,t} + \tilde{\eta}\varepsilon_{t+1}, \tag{10}$$

where Λ is an $n_2 \times n_2$ matrix, ε_{t+1} is an n_2 vector of i.i.d. shocks with zero mean and identity covariance matrix, and $\tilde{\eta}$ is an $n_2 \times n_2$ matrix that is introduced to allow for correlated structural shocks. We also assume that the eigenvalues of Λ are all less than one in absolute value, which guarantees that the exogenous stochastic process is stationary. The process (10) can be generalized to a nonlinear stochastic process of the form $x_{2,t+1} = H\left(x_{2,t}\right) + \tilde{\eta}\varepsilon_{t+1}$ for some nonlinear function H. But for our purposes, (10) is enough. The methods developed here can also be applied, with some modifications, when the exogenous stochastic process has a unit root.

We will solve the model using a first order perturbation (i.e. local linear approximation) around the non-stochastic steady state. To do this, it is convenient to add an "auxiliary" parameter (positive real number) σ that controls the "amount of uncertainty" in the model, and replace (10) by

$$x_{2,t+1} = \Lambda x_{2,t} + \sigma\tilde{\eta}\varepsilon_{t+1}. \tag{11}$$

When $\sigma = 0$ the model becomes deterministic while setting $\sigma = 1$ we recover the original model. The perturbation technique involves approximating the model around the deterministic model $\sigma = 0$.

Equation (9) together with the appropriate stability condition imply solutions in the form of policy functions $g : R^n \times R^+ \to R^m$ and $h : R^n \times R^+ \to R^n$ such that

$$y_t = g\left(x_t, \sigma\right) \tag{12}$$

$$x_{t+1} = h\left(x_t, \sigma\right) + \sigma\eta\varepsilon_{t+1} \tag{13}$$

[2] In the RBC example, x_t^1 is the stock of capital (a state variable that can be affected by agent choices) and x_t^2 is the level of technology (an exogenous state variable that cannot be affected by any agent).

where

$$\eta_{n\times n_2} = \begin{bmatrix} 0_{n_1\times n_2} \\ \tilde{\eta}_{n_2\times n_2} \end{bmatrix}.$$

Equations (12) and (13) are the policy functions of the model. In particular, (12) determines the control variables y_t as a function of the state variables x_t and (13) determines the evolution of the state variables. Under regularity conditions that will depend on the particular case being analized, the policy functions (12) and (13) are unique—more on this below.

We consider approximations around the non-stochastic steady state. The steady state is defined as the values of x and y that satisfy $x_{t+1} = x_t = \bar{x}$, $y_{t+1} = y_t = \bar{y}$, and that solve the system of equations

$$f(\bar{x}, \bar{y}, \bar{x}, \bar{y}) = 0. \tag{14}$$

This problem, in principle, has a solution as there are $n + m$ equations to find the $n + m$ unknowns.

The way we impose the appropriate stability conditions is to restrict attention to bounded equilibria in which the economy, absent any shock, converges to the steady state. As we will see below, this involves computing some eigenvalues and counting the number of stable versus unstable eigenvalues.

We now compute a first order Taylor approximation of the policy functions g and h around the point $(x, \sigma) = (\bar{x}, 0)$. Here, $\bar{x}$ is the steady state of the state variables of the deterministic economy and the scalar $\sigma = 0$ implies that we are perturbing the solution around an economy with no shocks. The Taylor expansions of the functions g and h around $(\bar{x}, 0)$ are

$$g(x, \sigma) \approx g(\bar{x}, 0) + g_x(\bar{x}, 0)(x - \bar{x}) + g_\sigma(\bar{x}, 0)\sigma \tag{15}$$

$$h(x, \sigma) \approx h(\bar{x}, 0) + h_x(\bar{x}, 0)(x - \bar{x}) + h_\sigma(\bar{x}, 0)\sigma \tag{16}$$

where g_x is the $(m\times n)$ Jacobian matrix of partial derivatives $\partial g_i/\partial x_j$ for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$; h_x is the $(n \times n)$ Jacobian with component $\partial h_i/\partial x_j$ for $i, j = 1, 2, ..., n$; g_σ is a $(m \times 1)$ vector with component $\partial g_i/\partial \sigma$ for $i = 1, 2, ..., m$; and h_σ is a $(n \times 1)$ vector with entries $\partial h_i/\partial \sigma$ for $i = 1, 2, ..., n$. All of these matrices are evaluated at the point $(x, \sigma) = (\bar{x}, 0)$.

By definition of the steady state, we know that

$$\begin{aligned} g(\bar{x}, 0) &= \bar{y} \\ h(\bar{x}, 0) &= \bar{x}. \end{aligned}$$

Therefore, the approximate policy functions are given by

$$y_t \approx \bar{y} + g_x(\bar{x}, 0)(x_t - \bar{x}) + g_\sigma(\bar{x}, 0)\sigma \tag{17}$$

$$x_{t+1} \approx \bar{x} + h_x(\bar{x}, 0)(x_t - \bar{x}) + h_\sigma(\bar{x}, 0)\sigma + \sigma\eta\varepsilon_{t+1} \tag{18}$$

Our task is to obtain the matrices $g_x(\bar{x},0)$, $g_\sigma(\bar{x},0)$, $h_x(\bar{x},0)$, and $h_\sigma(\bar{x},0)$. To that end, we start by substituting the true (unknown) policy functions (12) and (13) into (9), and define a new function $F: R^n \times R^+ \to R^{n+m}$ as

$$\begin{aligned} F(x_t,\sigma) &\equiv E_t\left[f(x_{t+1},y_{t+1},x_t,y_t)\right] \\ &= E_t\left[f\left(h(x_t,\sigma)+\sigma\eta\varepsilon_{t+1}, g\left(h(x_t,\sigma)+\sigma\eta\varepsilon_{t+1},\sigma\right), x_t, g(x_t,\sigma)\right)\right] = 0. \end{aligned} \tag{19}$$

Under the true solution, $F(x,\sigma)$ equals zero for all x and σ. This implies that all the derivatives of $F(x,\sigma)$ are also zero, or

$$F_\sigma(x,\sigma) = 0 \tag{20}$$

$$F_x(x,\sigma) = 0 \tag{21}$$

for all x and σ. Here, $F_\sigma(x,\sigma)$ is a vector of dimension $n+m$, and $F_x(x,\sigma)$ is a Jacobian matrix of dimension $(n+m)\times n$.

To find the matrices associated to the linear approximation—that is, $g_x(\bar{x},0)$, $g_\sigma(\bar{x},0)$, $h_x(\bar{x},0)$, and $h_\sigma(\bar{x},0)$—we note that (20) and (21) must also hold at the steady state $(\bar{x},0)$. We now use (19) to find the required derivatives. Differentiating (19) with respect to the scalar σ—and keeping the argument of the functions implicit to save on notation—gives

$$\begin{aligned} F_\sigma(\bar{x},0) &= E_t\left[f_{x'}(h_\sigma+\eta\varepsilon_{t+1}) + f_{y'}(g_x(h_\sigma+\eta\varepsilon_{t+1})+g_\sigma) + f_y g_\sigma\right] \\ &= f_{x'}(h_\sigma+\eta E_t[\varepsilon_{t+1}]) + f_{y'}(g_x(h_\sigma+\eta E_t[\varepsilon_{t+1}])+g_\sigma) + f_y g_\sigma \\ &= f_{x'}h_\sigma + f_{y'}(g_x h_\sigma + g_\sigma) + f_y g_\sigma \\ &= \left[f_{x'}+f_{y'}g_x\right]h_\sigma + \left[f_{y'}+f_y\right]g_\sigma, \end{aligned}$$

where the second equality uses that we are evaluating all expressions at the steady state, so that the only random object is the mean zero shock ε_{t+1}, and the third equality uses $E_t[\varepsilon_{t+1}]=0$. At this point we are after the derivatives $h_\sigma(\bar{x},0)$ and $g_\sigma(\bar{x},0)$. The expressions f_y, $f_{y'}$, and $f_{x'}$ are evaluated at the steady state and, thus, are known. As mentioned above, the previous expression has to be equal to zero, so that $F_\sigma(\bar{x},0)=0$ implies

$$\left[f_{x'}+f_{y'}g_x\right]h_\sigma + \left[f_{y'}+f_y\right]g_\sigma = \bar{0}_{(n+m)\times 1},$$

or

$$\begin{bmatrix} f_{x'}+f_{y'}g_x & f_{y'}+f_y \end{bmatrix}\begin{bmatrix} h_\sigma \\ g_\sigma \end{bmatrix} = \bar{0}.$$

This is a homogeneous linear sytem of equations in the unknowns (h_σ, g_σ). Clearly, $h_\sigma=0$ and $g_\sigma=0$ is *one solution* of this system. If there is another solution with $\tilde{h}_\sigma \neq 0$ or $\tilde{g}_\sigma \neq 0$, then $\alpha\tilde{h}_\sigma$ and $\alpha\tilde{g}_\sigma$

is also a solution of the system for any number α. But since we are looking for a pair of *unique* policy functions, it then must be the case that $h_\sigma = 0$ and $g_\sigma = 0$.

This has the important implication that the "amount of uncertainty", as reflected in the parameter σ, is irrelevant in a first order approximation to the policy functions. This result is usually called the "certainty equivalence principle," a property that no longer holds in higher order approximations of the model or using other solution methods. This also means that, to a first order approximation, the unconditional expected values of x_t and y_t are equal to their non-stochastic steady state values $\bar{x}$ and $\bar{y}$. In effect, to see that $E[x_t] = \bar{x}$ and $E[y_t] = \bar{y}$, note that using the linear approximation to the policy functions we obtain

$$\begin{aligned} E[x_{t+1}] &= E[h(x_t, \sigma)] \\ &\cong h(\bar{x}, 0) + h_x(\bar{x}, 0)(E[x_t] - \bar{x}) + h_\sigma(\bar{x}, 0)\sigma \\ &= \bar{x} + h_x(\bar{x}, 0)(E[x_t] - \bar{x}), \end{aligned}$$

where the third equality uses $h(\bar{x}, 0) = \bar{x}$ and $h_\sigma = 0$. Furthermore, using $E[x_t] = E[x_{t+1}]$ and rearranging gives

$$(I - h_x(\bar{x}, 0))(E[x_t] - \bar{x}) = 0.$$

If no eigenvalue of $h_x(\bar{x}, 0)$ is one in absolute value—as will be the case in the set of solutions we are looking for—then the only solution to this equation is $E[x_t] = \bar{x}$ which proves the claim.

We now discuss how to find the perturbation matrices $g_x(\bar{x}, 0)$ and $h_x(\bar{x}, 0)$. To do this, differentiate expression (19) with respect to x and evaluate the resulting expressions at the steady state $(\bar{x}, 0)$,

$$\begin{aligned} \bar{0}_{(n+m)\times n} &= F_x(\bar{x}, 0) \\ &= f_{x'} h_x + f_{y'} g_x h_x + f_x + f_y g_x \\ &= (f_{x'} + f_{y'} g_x) h_x + f_x + f_y g_x \\ &= \begin{bmatrix} f_{x'} & f_{y'} \end{bmatrix} \begin{bmatrix} I \\ g_x \end{bmatrix} h_x + \begin{bmatrix} f_x & f_y \end{bmatrix} \begin{bmatrix} I \\ g_x \end{bmatrix}. \end{aligned}$$

Rearranging gives

$$\begin{bmatrix} f_{x'} & f_{y'} \end{bmatrix} \begin{bmatrix} I \\ g_x \end{bmatrix} h_x = - \begin{bmatrix} f_x & f_y \end{bmatrix} \begin{bmatrix} I \\ g_x \end{bmatrix}. \tag{22}$$

This is a system of $(n+m) \times n$ quadratic equations in $(n+m) \times n$ unknowns given by the elements of g_x and h_x.

Let $A = \begin{bmatrix} f_{x'} & f_{y'} \end{bmatrix}$ and $B = - \begin{bmatrix} f_x & f_y \end{bmatrix}$ be $(n+m) \times (n+m)$ square matrices and note that, because $f_{x'}$, f_x, $f_{y'}$, and f_y are evaluated at the known steady state, A and B are known.

Let's define $\hat{x}_{t+j} = E_t[x_{t+j}] - \bar{x}$ and $\hat{y}_{t+j} = E_t[y_{t+j}] - \bar{y}$. When $j = 0$, we have $\hat{x}_t = x_t - \bar{x}$ and

$\hat{y}_t = y_t - \bar{y}$. Now post-multiply both sides of the system of equations (22) by $\hat{x}_t$ to obtain

$$A \begin{bmatrix} I \\ g_x \end{bmatrix} h_x \hat{x}_t = B \begin{bmatrix} I \\ g_x \end{bmatrix} \hat{x}_t$$

or

$$A \begin{bmatrix} h_x \hat{x}_t \\ g_x h_x \hat{x}_t \end{bmatrix} = B \begin{bmatrix} \hat{x}_t \\ g_x \hat{x}_t \end{bmatrix} \tag{23}$$

We now prove that $\hat{x}_{t+1} \approx h_x \hat{x}_t$, $\hat{y}_t \approx g_x \hat{x}_t$ and $\hat{y}_{t+1} \approx g_x h_x \hat{x}_t$. To see this, consider first

$$h_x \hat{x}_t = h_x \left(\bar{x}, 0\right) \left(x_t - \bar{x}\right).$$

Now, the Taylor expansion (18) and the result $h_\sigma = 0$, gives

$$x_{t+1} - \bar{x} \approx h_x \hat{x}_t + \eta \varepsilon_{t+1}. \tag{24}$$

Taking the conditional expectations at time t on both sides of the last equation then gives

$$E_t \left[x_{t+1}\right] - \bar{x} \equiv \hat{x}_{t+1} \approx h_x \hat{x}_t.$$

Likewise, using (17) and $g_\sigma = 0$ implies

$$y_t - \bar{y} \equiv \hat{y}_t \approx g_x \hat{x}_t$$

and

$$\hat{y}_{t+1} \approx g_x \hat{x}_{t+1} \approx g_x h_x \hat{x}_t.$$

Using these results into (23) gives

$$A \begin{bmatrix} \hat{x}_{t+1} \\ \hat{y}_{t+1} \end{bmatrix} = B \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \end{bmatrix}. \tag{25}$$

which is a usual representation of the equilibrium conditions of a linearized rational expectations model. Equivalently, this condition can be written as

$$A E_t \begin{bmatrix} x_{t+1} - \bar{x} \\ y_{t+1} - \bar{y} \end{bmatrix} = B \begin{bmatrix} x_t - \bar{x} \\ y_t - \bar{y} \end{bmatrix}, \tag{26}$$

We now use the QZ decomposition of Theorem 1 to solve this linear system of difference equations.

Define the vector

$$\hat{w}_t = \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \end{bmatrix} \tag{27}$$

and write the linearized system as

$$A\hat{w}_{t+1} = B\hat{w}_t \tag{28}$$

We look for non-explosive solutions that satisfy

$$\lim_{j\to\infty} \hat{w}_{t+j} = 0 \tag{29}$$

which means that, at any point in time, the system is expected to converge to the steady state.

By Theorem 1, there are unitary matrices Q and Z, and upper triangular matrices S and T such that

$$\begin{aligned} QAZ &= S \\ QBZ &= T. \end{aligned}$$

Moreover, the matrices can be arranged so that the diagonal elements s_{ii} and t_{ii} appear in any order. We choose the following ordering of the pairs (s_{ii}, t_{ii}) : the ones satisfying $|s_{ii}| > |t_{ii}|$ appear in the first block of diagonal elements of S and T. We call these pairs of elements the *stable generalized eigenvalues* for reasons that will become apparent below. We also impose the following assumption,

Assumption 1: There is no i such that $|s_{ii}| = |t_{ii}|$. In other words, the system (28) does not have a unit root.

Because Z is unitary (so that $ZZ^H = I$), we can premultiply (28) by Q to obtain

$$\begin{aligned} QA\hat{w}_{t+1} &= QB\hat{w}_t \\ QAZZ^H\hat{w}_{t+1} &= QBZZ^H\hat{w}_t \\ SZ^H\hat{w}_{t+1} &= TZ^H\hat{w}_t. \end{aligned}$$

where the second line uses the QZ decomposition.

Definining the new variable

$$z_t \equiv Z^H\hat{w}_t, \tag{30}$$

the system can be written as

$$Sz_{t+1} = Tz_t.$$

This is progress because we wrote the general first order difference equation as another first order difference equation but where the relevant matrices (S and T) are triangular.

Using that S and T are upper triangular matrices, we write the system in block form as

$$\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \begin{bmatrix} z_{t+1}^s \\ z_{t+1}^u \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} z_t^s \\ z_t^u \end{bmatrix}. \tag{31}$$

where the partition of the matrices is constructed so that S_{11} and T_{11} contain the pairs of elements s_{ii}, t_{ii} such that $|s_{ii}| > |t_{ii}|$, and z_t is partitioned accordingly—that is, we put the stable eigenvalues first in the system. This inequality together with Assumption 1 means that the diagonal elements of S_{22} and T_{22} satisfy $|s_{ii}| < |t_{ii}|$.

Now focus on the lower block of the system (31),

$$S_{22} z_{t+1}^u = T_{22} z_t^u,$$

Theorem 1 (Point 4) implies that there is no index i such that $s_{ii} = t_{ii} = 0$. Therefore, the inequality $|s_{ii}| < |t_{ii}|$ for the submatrices S_{22} and T_{22} implies that all the diagonal elements of T_{22} are non-zero. Furthermore, because T_{22} is upper triangular, Result 2 above implies that T_{22} is invertible. Therefore, premultiplying both sides of the previous expression by T_{22}^{-1} we obtain

$$T_{22}^{-1} S_{22} z_{t+1}^u = z_t^u.$$

Results 3 and 4 above imply that $T_{22}^{-1} S_{22}$ is also upper triangular with diagonal elements s_{ii}/t_{ii} (because the product of triangular matrices is triangular). Since the eigenvalues of an upper triangular matrix are its diagonal entries, it follows that the matrix $T_{22}^{-1} S_{22}$ *has all its eigenvalues smaller than one in absolute value*. It then follows that at least one element of z_t^u has to explode to infinity in absolute value *unless* $z_t^u = 0$ for all t.[3] In other words, the only stable solution of the lower block of the system (31) is $z_t^u = 0$ for all t.

We now proceed to find the solution of the first block of equations in the system (31) (that corresponding to z_t^s). Using that $z_t^u = 0$ for all t, the first block of (31) implies

$$S_{11} z_{t+1}^s = T_{11} z_t^s.$$

Now, because S_{11} and T_{11} are such that $|s_{ii}| > |t_{ii}|$, it follows that all the diagonal elements of S_{11} are different from zero and, therefore, Result 2 implies that S_{11} is invertible. We thus have

$$z_{t+1}^s = S_{11}^{-1} T_{11} z_t^s. \tag{32}$$

By Results 3 and 4, $S_{11}^{-1} T_{11}$ is upper triangular with diagonal elements $|t_{ii}/s_{ii}| < 1$ for all i (this follows

[3] To obtain intuition, note that if S_{22} were invertible (which may not be), then $\left(T_{22}^{-1} S_{22}\right)^{-1} = S_{22}^{-1} T_{22}$ would be an upper triangular matrix with all its eigenvalues greater than 1 in absolute value and $z_{t+1}^u = S_{22}^{-1} T_{22} z_t^u$. This means that, unless $z_t^u = 0$, z_t^u would explode towards (plus or minus) infinity.

from $|s_{ii}| > |t_{ii}|$ on S_{11} and T_{11}). Therefore, (32) is a stable first order difference equation that converges to zero as $t \to \infty$ for any value of z_0^s. In other words, (32) is the solution to the first block of the system (31).

However, we are interested in the solution in terms of the variables $\hat{w}_t$, not in terms of z_t. To recover the solution in terms of $\hat{w}_t$, we use that Z^H is invertible (with inverse Z) and, from (30), it follows that

$$\hat{w}_t \equiv \begin{bmatrix} \hat{w}_t^s \\ \hat{w}_t^u \end{bmatrix} = Z \begin{bmatrix} z_t^s \\ z_t^u \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} z_t^s \\ z_t^u \end{bmatrix} \tag{33}$$

where $\hat{w}_t$ and Z were partitioned in terms of the stable and unstable eigenvalues as we did above with S and T. That is, $\hat{w}_t$ is a linear combination of z_t^s and z_t^u. But because $z_t^u = 0$ for all t and $z_t^s \to 0$ as $t \to \infty$ for any initial condition, it follows that $\hat{w}_t$ also converges to zero as $t \to \infty$. This proves that the stability condition (29) is satisfied.

Consider now the first block of (33), where we use $z_t^u = 0$ for all t,

$$\hat{w}_t^s = Z_{11} z_t^s.$$

We now impose the following assumption:

Assumption 2: Z_{11} is invertible.

With this assumption,

$$z_t^s = Z_{11}^{-1} \hat{w}_t^s.$$

Combining this expression with (32) gives

$$Z_{11}^{-1} \hat{w}_{t+1}^s = S_{11}^{-1} T_{11} Z_{11}^{-1} \hat{w}_t^s.$$

Now, pre-multiplying both sides of this expression by Z_{11} gives

$$\hat{w}_{t+1}^s = H \hat{w}_t^s. \tag{34}$$

where

$$H = Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1}.$$

Under Assumption 2, the eigenvalues of H are the same as the eigenvalues of $S_{11}^{-1} T_{11}$,[4] which is an upper triangular matrix with eigenvalues $|t_{ii}/s_{ii}| < 1$. This is the solution to the first block of equations $\hat{w}_t^s$.

[4] This follow because of the following result: Let A by a square $n \times n$ matrix and G an $n \times n$ non-singular matrix. Then, A and GAG^{-1} have the same eigenvalues. Proof:

$$\det\left(GAG^{-1} - \lambda I\right) = \det\left(GAG^{-1} - \lambda GG^{-1}\right) = \det\left(G\left(A - \lambda I\right)G^{-1}\right) = \det\left(G\right)\det\left(A - \lambda I\right)\det\left(G^{-1}\right) = \det\left(A - \lambda I\right)$$

because $\det\left(G^{-1}\right) = 1/\det\left(G\right)$.

We now obtain the solution of $\hat{w}_t^u$. The second block of (33) together with $z_t^s = Z_{11}^{-1}\hat{w}_t^s$ implies

$$\hat{w}_t^u = Z_{21} z_t^s = Z_{21} Z_{11}^{-1} \hat{w}_t^s, \tag{35}$$

which gives the solution of $\hat{w}_t^u$ as a function of $\hat{w}_t^s$.

Summarizing the results so far, given an initial condition for w_t^s we have the solution for the entire vector $\hat{w}_t$. As we will show below, in a locally unique rational expectation equilibrium w_t^s coincides with the vector of state variables.

Local existence and uniqueness of the equilibrium

We were able to find a solution to the linear stochastic difference equation (28). However, we didn't find yet the coefficient matrices $h_x(\bar{x}, 0)$ and $g_x(\bar{x}, 0)$ of the linear approximation to the policy functions. This is what we do now. In the process, we analyze the important issues of local existence and uniqueness of the equilibrium.

Local uniqueness of equilibrium and the Blanchard-Kahn condition.—

The **Blanchard and Kahn condition** is the requirement that *the number of stable generalized eigenvalues of the matrix pair* (A, B) *(that is, the number of elements* i *such that* $|t_{ii}/s_{ii}| < 1$*) is exactly equal to the number of state variables* n*.*

If the Blanchard and Kahn condition is satisfied, then the equilibrium of the DSGE model exists and is locally unique. Indeed, if the Blanchard and Kahn condition is satisfied, then Z_{11} is of size $n \times n$ and $\hat{w}_t^s$ is of size $n \times 1$. The definition of $\hat{w}_t$ in equation (27) and the last observation imply that

$$\begin{aligned} \hat{w}_{t+1}^s &= \hat{x}_{t+1} = E_t[x_{t+1}] - \bar{x} \\ \hat{w}_t^s &= \hat{x}_t = x_t - \bar{x}. \end{aligned}$$

Then, (34) implies

$$E_t[x_{t+1}] - \bar{x} = H(x_t - \bar{x}),$$

where $H = Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1}$. Dropping the expectation operator and using (24) leads to

$$x_{t+1} = \bar{x} + H(x_t - \bar{x}) + \eta \varepsilon_{t+1}. \tag{36}$$

This implies that $h_x(\bar{x}, 0) = H = Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1}$ is the matrix of elements that we were looking for. Since x_t is uniquely determined from $\hat{w}_t^s$, the linearized policy function is also unique.

Next, using that $\hat{w}_t^u = y_t - \bar{y}$ (because $\hat{w}_t^s = \hat{x}_t$), equation (35) implies

$$y_t = \bar{y} + Z_{21} Z_{11}^{-1}(x_t - \bar{x}), \tag{37}$$

so that $g_x(\bar{x}, 0) = G = Z_{21}Z_{11}^{-1}$. Again, since y_t is uniquely determined from x_t, the policy function for control variables is also locally unique.

Summarizing, we showed that if the number of stable generalized eigenvalues with absolute value less than one is equal to the number of predetermined variables x_t (the so-called Blanchard and Kahn condition), the equilibrium exists locally around the steady state and is unique. In some models, like the RBC model described above, it can be shown that the Blanchard and Kahn condition must hold and that the equilibrium is locally unique.

No local existence of the equilibrium.—

Suppose now that the number of generalized eigenvalues of (A, B) with absolute value less than unity ($|t_{ii}/s_{ii}| < 1$) is smaller than the number of state variables n. Specifically, suppose that Z_{11} is of size $(n-q) \times (n-q)$ for $0 < q < n$. In this case, $\hat{w}_t^s$ has less elements than the state vector $\hat{x}_t$. The vectors $\hat{w}_t^s$ and $\hat{w}_t^u$ take the form

$$\hat{w}_t^s = \hat{x}_t^a \text{ and } \hat{w}_t^u = \begin{bmatrix} \hat{x}_t^b \\ \hat{y}_t \end{bmatrix},$$

where

$$\hat{x}_t = \begin{bmatrix} \hat{x}_t^a \\ \hat{x}_t^b \end{bmatrix}.$$

In this case, the solution of the difference equation (34) and (35) imply

$$\hat{x}_{t+1}^a = H\hat{x}_t^a$$

$$\begin{bmatrix} \hat{x}_t^b \\ \hat{y}_t \end{bmatrix} = G\hat{x}_t^a.$$

The last expression states that $\hat{x}_t^b$ is determined by $\hat{x}_t^a$. But this is impossible because $\hat{x}_t^b$ is a predetermined variable independent of $\hat{x}_t^a$. Therefore, there is no local equilibrium when the number of generalized eigenvalues of (A, B) with absolute value less than unity is smaller than the number of state variables n.

Local indeterminacy of the equilibrium.—

Finally, suppose that the number of generalized eigenvalues of (A, B) with absolute value less than one is greater than the number of state variables, n. Specifically, suppose that there are $n + q$, for $0 < q \leq m$ generalized eigenvalues with absolute value less than one. Then Z_{11} is size $(n+q) \times (n+q)$. Thus, $\hat{w}_t^s$ has more elements than $\hat{x}_t$ and $\hat{w}_t^u$ has less elements than $\hat{y}_t$. In particular,

$$\hat{w}_t^s = \begin{bmatrix} \hat{x}_t \\ \hat{y}_t^a \end{bmatrix}; \ \hat{w}_t^u = \hat{y}_t^b$$

where $\hat{y}_t^a$ is a vector with the first q elements of the vector $\hat{y}_t$, and $\hat{y}_t^b$ is a vector with the remaning $m-q$ elements of $\hat{y}_t$, so that

$$\hat{y}_t = \begin{bmatrix} \hat{y}_t^a \\ \hat{y}_t^b \end{bmatrix}.$$

According to (34) and (35), the laws of motion for $\hat{x}_t$ and $\hat{y}_t$ are as follows

$$\begin{bmatrix} \hat{x}_{t+1} \\ \hat{y}_{t+1}^a \end{bmatrix} = H \begin{bmatrix} \hat{x}_t \\ \hat{y}_t^a \end{bmatrix} \tag{38}$$

$$\hat{y}_t^b = G \begin{bmatrix} \hat{x}_t \\ \hat{y}_t^a \end{bmatrix}. \tag{39}$$

Now, since $\hat{y}_t^a$ is a jump variable, it is not predetermined at time t. Therefore, we can freely pick $\hat{y}_t^a$ at an arbitrary period, say $t=0$, and then solve the above system in the variables $\hat{x}_t$ and $\hat{y}_t$. But since $\hat{y}_0^a$ was arbitrary, this means that the equilibrium is indeterminate.

Furthermore, there is even a deeper sense in which the equilibrium is indeterminate. The above equations are written in terms of variables with expected values. Since there is nothing that ties $\hat{y}_t^a$ to previous decisions, we can always drop the expectation operator and write the above system as

$$\begin{bmatrix} x_{t+1} \\ y_{t+1}^a \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{y}^a \end{bmatrix} + H \begin{bmatrix} x_t - \bar{x} \\ y_t^a - \bar{y}^a \end{bmatrix} + \begin{bmatrix} \eta & 0 \\ \nu_\varepsilon & \nu_\mu \end{bmatrix} \begin{bmatrix} \varepsilon_{t+1} \\ \mu_{t+1} \end{bmatrix}$$

$$y_t^b - \bar{y}^b = G \begin{bmatrix} x_t - \bar{x} \\ y_t^a - \bar{y}^a \end{bmatrix},$$

where μ_{t+1} is an arbitrary mean zero stochastic process of size $q \times 1$ and variance covariance matrix equal to the identity matrix. The matrices ν_ε and ν_μ are also arbitrary. To see that this solves the difference equation, take conditional expectations and we return to the system (38) and (39).

Summarizing, this shows that if the number of generalized eigenvalues of (A, B) with absolute value less than one is greater than the number of state variables, we can construct many equilibria where non-fundamental uncertainty or sunspots, μ_{t+1}, affect the evolution of the equilibrium quantities of the model. There is a literature that exploits this indeterminacy to study sunspots and multiple equilibria.

UNCONDITIONAL SECOND MOMENTS

This part of the note discusses how to compute unconditional second moments using the first order approximation to the policy functions. Because the policy functions end up being linear, it is easy to compute unconditional (population) second moments for all the variables of interest.

We first compute the matrices $h_x(\bar{x}, 0) \equiv h_x$ and $g_x(\bar{x}, 0) \equiv g_x$ using the method described above. If

we let $\tilde{x}_t \equiv x_t - \bar{x}$ and $\tilde{y}_t = y_t - \bar{y}$ denote the deviation of the variables x_t and y_t relative to their steady state values, we can write the linearized policy functions as

$$\tilde{x}_{t+1} = h_x \tilde{x}_t + \eta \varepsilon_{t+1} \tag{40}$$

$$\tilde{y}_t = g_x \tilde{x}_t. \tag{41}$$

where we are evaluating the policy function at the original economy (that is, when $\sigma = 1$) Our objective in this section is to compute all the second moments of the variables $\tilde{x}_t$ and $\tilde{y}_t$.

Covariance matrix of x_t

Let

$$\Sigma_x \equiv E\left[(x_t - \bar{x})(x_t - \bar{x})'\right] = E\left[\tilde{x}_t \tilde{x}_t'\right]$$

Now take equation (40) and post-multiply both sides by its transpose:

$$\begin{aligned}
\tilde{x}_{t+1}\tilde{x}_{t+1}' &= (h_x \tilde{x}_t + \eta \varepsilon_{t+1})(h_x \tilde{x}_t + \eta \varepsilon_{t+1})' \\
&= (h_x \tilde{x}_t + \eta \varepsilon_{t+1})\left(\tilde{x}_t' h_x' + \varepsilon_{t+1}' \eta'\right) \\
&= h_x\left(\tilde{x}_t \tilde{x}_t'\right) h_x' + \eta\left(\varepsilon_{t+1}\tilde{x}_t'\right) h_x' + h_x\left(\tilde{x}_t \varepsilon_{t+1}'\right)\eta' + \eta \varepsilon_{t+1}\varepsilon_{t+1}'\eta'
\end{aligned}$$

Taking uncondicional expectations and using $E\left(\varepsilon_{t+1}\tilde{x}_t'\right) = E\left(\tilde{x}_t \varepsilon_{t+1}'\right) = 0$ because ε_{t+1} is an i.i.d. shock we obtain

$$E\left[\tilde{x}_{t+1}\tilde{x}_{t+1}'\right] = h_x E\left[\tilde{x}_t \tilde{x}_t'\right] h_x' + \eta E\left[\varepsilon_{t+1}\varepsilon_{t+1}'\right]\eta'$$

Now, using that $\tilde{x}_t$ is a covariance stationary process, so that $E\left[\tilde{x}_{t+1}\tilde{x}_{t+1}'\right] = E\left[\tilde{x}_t \tilde{x}_t'\right] = \Sigma_x$ and that $E\left[\varepsilon_{t+1}\varepsilon_{t+1}'\right] = I$, we have that Σ_x satisfies the following equation

$$\Sigma_x = h_x \Sigma_x h_x' + \Sigma_\varepsilon. \tag{42}$$

where we define $\Sigma_\varepsilon \equiv \eta\eta'$. This is an equation for Σ_x. The following are two methods to compute Σ_x.

Method 1: This method uses the following useful result. Let A, B, and C be matrices such that ABC is well defined. Then,

$$vec(ABC) = \left(C' \otimes A\right) vec(B),$$

where $vec(A)$ is the "vectorization" operator that transforms a matrix A into a column vector by stacking the column vectors of A, and $A \otimes B$ denotes the Kronecker product of the matrices A and B.

Now apply the *vec* operator to both sides of (42),

$$vec(\Sigma_x) = vec\left(h_x \Sigma_x h_x'\right) + vec(\Sigma_\varepsilon)$$

Using the result described above

$$vec\left(\Sigma_x\right) = h_x \otimes h_x vec\left(\Sigma_x\right) + vec(\Sigma_\varepsilon)$$

Therefore, the value of Σ_x solves

$$vec\left(\Sigma_x\right) = \left(I - h_x \otimes h_x\right)^{-1} vec(\Sigma_\varepsilon).$$

One possible drawback of this approach is that we have to invert a matrix of size $n^2 \times n^2$.

Method 2: The second method uses an iterative algorithm called *doubling algorithm.* We start with a guess of Σ_x, which we call Σ_x^0, and compute iteratively updated guesses of Σ_x according to the following algorithm: for $j = 1, 2, ...$ compute

$$\Sigma_x^j = h_x \Sigma_x^{j-1} h_x' + \Sigma_\varepsilon.$$

Stop the algorithm when $\Sigma_x^j \approx \Sigma_x^{j-1}$. We can start the algorithm with $\Sigma_x^0 = I$.

Autocovariances of x_t

We now can easily compute the autocovariances of x_t of any order. Consider the autocovariance of order $\tau > 0$, which we denote by $\Sigma_x(\tau)$. Then,

$$\Sigma_x(\tau) = E\left[\tilde{x}_t \tilde{x}_{t-\tau}'\right].$$

Note also that $\Sigma_x(0) = \Sigma_x$ is the covariance matrix of x_t.

To compute this matrix, let $\mu_t \equiv \eta \varepsilon_t$, and write

$$\begin{aligned}
\tilde{x}_t &= h_x \tilde{x}_{t-1} + \mu_t \\
&= h_x\left(h_x \tilde{x}_{t-2} + \mu_{t-1}\right) + \mu_t \\
&= h_x^2 \tilde{x}_{t-2} + h_x \mu_{t-1} + \mu_t \\
&= h_x^2\left(h_x \tilde{x}_{t-3} + \mu_{t-2}\right) + h_x \mu_{t-1} + \mu_t \\
&= h_x^3 \tilde{x}_{t-3} + h_x^2 \mu_{t-2} + h_x \mu_{t-1} + \mu_t \\
&\vdots \\
&= h_x^\tau \tilde{x}_{t-\tau} + h_x^{\tau-1} \mu_{t-(\tau-1)} + h_x^{\tau-2} \mu_{t-(\tau-2)} + ... + h_x \mu_{t-1} + \mu_t
\end{aligned}$$

or

$$\tilde{x}_t = h_x^\tau \tilde{x}_{t-\tau} + \sum_{j=0}^{\tau-1} h_x^j \mu_{t-j}$$

Therefore

$$\begin{aligned}
\Sigma_x(\tau) &= E\left[\tilde{x}_t\tilde{x}'_{t-\tau}\right] \\
&= E\left[\left(h_x^\tau\tilde{x}_{t-\tau}+\sum_{j=0}^{\tau-1}h_x^j\mu_{t-j}\right)\tilde{x}'_{t-\tau}\right] \\
&= E\left[h_x^\tau\tilde{x}_{t-\tau}\tilde{x}'_{t-\tau}+\sum_{j=0}^{\tau-1}h_x^j\mu_{t-j}\tilde{x}'_{t-\tau}\right] \\
&= h_x^\tau E\left[\tilde{x}_{t-\tau}\tilde{x}'_{t-\tau}\right]+\sum_{j=0}^{\tau-1}h_x^j E\left[\mu_{t-j}\tilde{x}'_{t-\tau}\right] \\
&= h_x^\tau\Sigma_x(0)
\end{aligned}$$

where we used that $E\left[\mu_{t-j}\tilde{x}'_{t-\tau}\right]=E\left[\eta\varepsilon_{t-j}\tilde{x}'_{t-\tau}\right]=0$ for all j and τ.

Summarizing, the autocovariance matrices of x_t of order τ satisfy the equation

$$\Sigma_x(\tau)=h_x^\tau\Sigma_x. \tag{43}$$

Equivalently, we can compute iteratively this matrices as $\Sigma_x(\tau)=h_x\Sigma_x(\tau-1)$ starting with $\tau=1$.

Second moments of the control variables y_t

We now compute the second moments of y_t using the second moments of x_t. To that end, let

$$\Sigma_y=E\left[\tilde{y}_t\tilde{y}'_t\right]$$

$$\Sigma_y(\tau)=E\left[\tilde{y}_t\tilde{y}'_{t-\tau}\right]$$

Note that (41) implies

$$\begin{aligned}
\Sigma_y &= E\left[\tilde{y}_t\tilde{y}'_t\right] \\
&= E\left[\left(g_x\tilde{x}_t\right)\left(g_x\tilde{x}_t\right)'\right] \\
&= g_xE\left[\tilde{x}_t\tilde{x}'_t\right]g'_x
\end{aligned}$$

or

$$\Sigma_y=g_x\Sigma_xg'_x. \tag{44}$$

Likewise,

$$\begin{aligned}
\Sigma_y(\tau) &= E\left[\tilde{y}_t\tilde{y}'_{t-\tau}\right] \\
&= g_xE\left[\tilde{x}_t\tilde{x}'_{t-\tau}\right]g'_x
\end{aligned}$$

so that

$$\Sigma_y(\tau) = g_x \Sigma_x(\tau) g_x'. \tag{45}$$

Spectral density of x_t and y_t

Write the evolution of $\tilde{x}_t$ in terms of the lag operator L,

$$(1 - h_x L)\tilde{x}_t = \eta \varepsilon_t$$

If the model has no unit roots, then all the eigenvalues of the matrix $h_x(\bar{x}, 0)$ are less than one in absolute value. We can therefore obtain the infinite order moving average representation of $\tilde{x}_t$ inverting the lag-polynomial

$$\tilde{x}_t = (I - h_x L)^{-1} \eta \varepsilon_t,$$

where I an $n \times n$ identity matrix. Using results from our second lecture (or see formula 10.4.43 in page 277 of Hamilton, 1994), we can compute the spectral density of $\tilde{x}_t$ using the spectral density of $\eta \varepsilon_t$,

$$S_{\tilde{x}}(\omega) = \left(I - h_x e^{-i\omega}\right)^{-1} S_{\eta\varepsilon_t}(\omega) \left[\left(I - h_x e^{-i\omega}\right)^{-1}\right]^*$$

Using properties of the inverse and the conjugate transpose gives

$$S_{\tilde{x}}(\omega) = \left(I - h_x e^{-i\omega}\right)^{-1} S_{\eta\varepsilon_t}(\omega) \left(I - h_x' e^{i\omega}\right)^{-1}$$

But using that ε_t is uncorrelated with identity covariance matrix, we obtain

$$S_{\eta\varepsilon_t}(\omega) = \Sigma_\varepsilon (= \eta\eta').$$

Therefore, the spectral density of the state variables $\tilde{x}_t$ is

$$S_{\tilde{x}}(\omega) = \left(I - h_x e^{-i\omega}\right)^{-1} \Sigma_\varepsilon \left(I - h_x' e^{i\omega}\right)^{-1}. \tag{46}$$

To obtain the spectral density matrix of the control variables $\tilde{y}_t$, we note that

$$\tilde{y}_t = g_x \tilde{x}_t.$$

It then follows that

$$S_{\tilde{y}}(\omega) = g_x S_{\tilde{x}}(\omega) g_x'$$

or

$$S_{\tilde{y}}(\omega) = g_x \left(I - h_x e^{-i\omega}\right)^{-1} \Sigma_\varepsilon \left(I - h_x' e^{i\omega}\right)^{-1} g_x'. \tag{47}$$

IMPULSE RESPONSE FUNCTIONS (IR)

The easiest way to compute an impulse response function is to simulate the model for a particular history of shocks to the variable of interest in the shock vector ε_t. Suppose that we want to compute the impulse responses to a shock of one standard deviation in the first element of the vector ε_t, denoted by $\varepsilon_{1,t}$, at time $t = 0$. We simply simulate the model by setting $\varepsilon_{1,t} = 1$ for $t = 0$ and $\varepsilon_{1,t} = 0$ for all $t > 0$, and set all the other shocks to zero $\varepsilon_{j,t} = 0$ for all t and all $j \neq 1$.

But we can also compute the impulse response analytically using the policy functions of the model. We use the following definition

The impulse response to a variable z_t in period $t + j$ to an arbitrary impulse in period t (i.e. an arbitrary shock to the vector ε_t) is defined as

$$IR\left(z_{t+j}\right) = E_t\left[z_{t+j}\right] - E_{t-1}\left[z_{t+j}\right].$$

Notice the timing of the expectation operators. The impulse response tells us the new information that we acquire exactly at time t of the variable z_t in period $t + j$. New information means what we know at time t that we didn't know at time $t - 1$.

We start by computing the impulse response to the state variables x_t. Using the policy function $\tilde{x}_{t+1} = h_x\tilde{x}_t + \eta\varepsilon_{t+1}$ we have $E_t\left[\tilde{x}_{t+j}\right] = h_x^j\tilde{x}_t$. Suppose that at time $t = 0$ there is an initial impulse $\tilde{x}_0 = \eta\varepsilon_0$ (here we use $\tilde{x}_{-1} = 0$ so that the economy starts at the steady state). Applying the law of iterated expectations and $E_{-1}\varepsilon_0 = 0$ we have

$$\begin{aligned} E_0\left[\tilde{x}_t\right] &= h_x^t\tilde{x}_0 \\ E_{-1}\left[\tilde{x}_t\right] &= h_x^tE_{-1}\tilde{x}_0 = h_x^t\eta E_{-1}\varepsilon_0 = 0. \end{aligned}$$

Therefore, the impulse response to x_t at time t to an impulse ε_0 at time 0 is

$$IR\left(\tilde{x}_t\right) = E_0\left[\tilde{x}_t\right] - E_{-1}\left[\tilde{x}_t\right] = h_x^t\tilde{x}_0.$$

The impulse to the vector of control variables $\tilde{y}_t = g_x\tilde{x}_t$ is thus

$$IR\left(\tilde{y}_t\right) = g_xh_x^t\tilde{x}_0.$$

REFERENCES

[1] Gomme, P. and P. Klein (2011). Second-order approximation of dynamic models without the use of tensors. *Journal of Economic Dynamics and Control* 35, pp. 604-615.

[2] Klein, Paul (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control* 24, pp. 1405-1423.

[3] Schmitt-Grohé, S. and M. Uribe (2017). Open Economy Macroeconomics. Princeton University Press. Material here: http://www.columbia.edu/~mu2166/book/

BASIC RBC MODEL

This note shows how to solve, step-by-step, a plain vanilla real business cycle (RBC) model using a log-linearization of the equilibrium conditions around the steady state. We will use Paul Klein's matlab function `solab.m`. This program uses the Generalized Schur decomposition as described in a previous note to solve the first order stochastic difference equation derived from the log-linearized model. The general procedure to solve a model consists of the following steps:

1. Find the equilibrium conditions of the model.
2. Calibrate and find the steady state of the model.
3. Log-linearize the equilibrium conditions of the model around the steady state.
4. Write the linearized system of difference equations in a format that can be used in the Matlab function `solab.m`. This requires writing the system of equations as

$$\mathbf{A}E_t\left[\mathbf{z}_{t+1}\right] = \mathbf{B}\mathbf{z}_t$$

   where $\mathbf{z}_t$ is a vector that contains all the variables in the model ordered in a particular way (see below), and $\mathbf{A}$ and $\mathbf{B}$ are square matrices conformable with the vector $\mathbf{z}_t$.
5. Call the `solab.m` routine to find the approximate policy functions. Then compute impulse responses, simulations, second moments, and so forth.

We consider a standard RBC model cast as the following planner's problem

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \left[\log c_t - \eta \frac{l_t^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}}\right]$$

subject to

$$c_t + k_{t+1} = A_t k_t^{\alpha} l_t^{1-\alpha} + (1-\delta)\, k_t$$

given initial conditions A_0, k_0, and a law of motion for the technology process that we specify below. The parameter η is a constant affecting the disutility of working and ν is the Frisch elasticity of labor supply.

Let λ_t denote the multiplier on the constraint and write the Lagrangian

$$L = E_0 \sum_{t=0}^{\infty} \beta^t \left[\log c_t - \eta \frac{l_t^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}} - \lambda_t \left[c_t + k_{t+1} - A_t k_t^{\alpha} l_t^{1-\alpha} - (1-\delta)\, k_t\right]\right]$$

The first order conditions with respect to c_t, l_t, and k_{t+1} are, respectively,

$$\frac{1}{c_t} = \lambda_t$$

$$\eta l_t^{\frac{1}{\nu}} = \lambda_t (1-\alpha) A_t k_t^{\alpha} l_t^{-\alpha}$$

$$\lambda_t = \beta E_t \left[\lambda_{t+1} \left(\alpha A_{t+1} k_{t+1}^{\alpha-1} l_{t+1}^{1-\alpha} + 1 - \delta \right) \right]$$

plus the feasibility constraint. The transversality condition of this problem is

$$\lim_{T \to \infty} E_0 \left[\beta^T \lambda_T k_{T+1} \right] = 0.$$

Shocks. The logarithm of TFP follows an AR(1) process

$$\log A_{t+1} = \rho \log A_t + \varepsilon_{t+1}$$

where ε_{t+1} is i.i.d. normal with mean 0 and variance σ_{ε}^2.

Equilibrium equations

As written, the control variables of the model are consumption c_t, labor l_t, and the multiplier λ_t. But we are also interested in output (y_t) and investment (x_t). We will augment the above system to include a few equations that will allow us to obtain the policy functions for output and investment in a direct fashion. In particular, recall that output is given by

$$y_t = A_t k_t^{\alpha} l_t^{1-\alpha}$$

and investment is

$$x_t = k_{t+1} - (1-\delta) k_t$$

Then, we can write the marginal product of capital and labor as

$$\begin{aligned} (1-\alpha) A_t k_t^{\alpha} l_t^{-\alpha} &= (1-\alpha) y_t / l_t \text{ and} \\ \alpha A_t k_t^{\alpha-1} l_t^{1-\alpha} &= \alpha y_t / k_t. \end{aligned}$$

It then follows that we can write the equilibrium conditions as the following system of 7 equations

$$\frac{1}{c_t} = \lambda_t \tag{1}$$

$$\eta l_t^{\frac{1}{\nu}} = \lambda_t (1-\alpha) \frac{y_t}{l_t} \tag{2}$$

$$\lambda_t = \beta E_t \left[\lambda_{t+1} \left(\alpha \frac{y_{t+1}}{k_{t+1}} + 1 - \delta \right) \right] \tag{3}$$

$$y_t = A_t k_t^{\alpha} l_t^{1-\alpha} \tag{4}$$

$$c_t + x_t = y_t \tag{5}$$

$$x_t = k_{t+1} - (1-\delta) k_t \tag{6}$$

$$\log A_{t+1} = \rho \log A_t + \varepsilon_{t+1} \tag{7}$$

Steady state and calibration

The second step in the procedure consists of finding the non-stochastic steady steate of the economy and calibrating the model. In steady state the system (1)-(7) becomes

$$\frac{1}{\bar{c}} = \bar{\lambda} \tag{8}$$

$$\eta \bar{l}^{\frac{1}{\nu}} = \bar{\lambda} (1-\alpha) \bar{y}/\bar{l} \tag{9}$$

$$1 = \beta \left(\alpha \bar{y}/\bar{k} + 1 - \delta \right) \tag{10}$$

$$\bar{y} = \bar{A} \bar{k}^{\alpha} \bar{l}^{1-\alpha} \tag{11}$$

$$\bar{c} + \bar{x} = \bar{y} \tag{12}$$

$$\bar{x} = \delta \bar{k} \tag{13}$$

$$\bar{A} = 1. \tag{14}$$

The steady state quantities satisfy the previous system of equations. One possibility is to solve the system as a function of the parameters of the model. It is easier, however, to perform a calibration that simultaneously sets the parameters of the model and delivers the steady state quantities. The following is one possibility:

Calibration: We set some numbers to match certain features of the data. For example, the parameter α is the capital share in output. NIPA accounts for the U.S. imply a capital share of about $\alpha = 1/3$. Second, we calibrate the a long-run (gross) real interest rate of $\bar{R} = 0.01$ (1% per quarter). In the model, the gross real interest rate in steady state satisfies

$$\alpha \frac{\bar{y}}{\bar{k}} + 1 - \delta = \bar{R}. \tag{15}$$

Given the value for α chosen before, this equation imposes a relation between the steady state

output to capital ratio $\left(\bar{y}/\bar{k}\right)$ and the depreciation rate δ. Equation (10) then implies that β must satisfy

$$\frac{1}{\beta} = \bar{R} \rightarrow \beta = \frac{1}{1.01} \approx 0.99.$$

Next, we want to calibrate the model to match some average (long-run) investment rate $\bar{x}/\bar{y}$. Equation (13) can be written as

$$\frac{\bar{x}}{\bar{y}} = \delta \frac{\bar{k}}{\bar{y}}$$

But using the calibration relation (15) we can write the last equation as

$$\frac{\bar{x}}{\bar{y}} = \delta \left(\frac{\alpha}{\bar{R} - (1 - \delta)} \right)$$

We can use this equation to solve for the value of δ :

$$\delta = \frac{\left(\bar{R} - 1\right)\left(\bar{x}/\bar{y}\right)}{\alpha - \left(\bar{x}/\bar{y}\right)}.$$

Given a target value for the investment rate $\bar{x}/\bar{y} = 0.21$ (this is roughly the value for the US for a broad concept of investment that includes durable consumption goods), and the values $\bar{R} = 1.01$ and $\alpha = 1/3$ we obtain

$$\delta = \frac{0.01 \times 0.21}{0.33 - 0.21} \approx 0.017.$$

This implies an annualized depreciation rate of about 6.6%:

$$(1 - 0.017)^4 = 1 - \delta^{annual} \Rightarrow \delta^{annual} = 1 - (1 - 0.017)^4 \cong 0.0663.$$

(In any case, in the model we used the quarterly depreciation rate.)

Next, we calibrate the model to deliver an average labor input value of $\bar{l} = 1/3$. This is roughly the fraction of total weekly hours that workers spend working. Using (14), we can write (11) as

$$\bar{y} = \bar{k}^{\alpha} \bar{l}^{1-\alpha}.$$

Dividing by $\bar{k}$

$$\frac{\bar{y}}{\bar{k}} = \left(\frac{\bar{l}}{\bar{k}} \right)^{1-\alpha}$$

and using the calibration condition (15) we can write

$$\left(\frac{\bar{R}-(1-\delta)}{\alpha}\right)=\left(\frac{\bar{l}}{\bar{k}}\right)^{1-\alpha}$$

or

$$\bar{k}=\bar{l}\left(\frac{\alpha}{\bar{R}-(1-\delta)}\right)^{\frac{1}{1-\alpha}} \tag{16}$$

Given the target value of $\bar{l}=1/3$ and the other parameters already calibrated, this equation delivers the steady state level of capital. For the parameter values described above, we find

$$\bar{k}=\frac{1}{3}\left(\frac{1/3}{1.01-(1-0.017)}\right)^{\frac{1}{1-1/3}}\approx 14.46.$$

The steady state level of output is thus

$$\bar{y}=\bar{A}\bar{k}^{\alpha}\bar{l}^{1-\alpha}\approx 1.17.$$

The steady state level of consumption follows from the feasibility condition (12)

$$\bar{c}=\bar{y}-\bar{x}=\bar{y}\left(1-\frac{\bar{x}}{\bar{y}}\right)=1.17\,(1-0.21)\approx 0.93.$$

Given $\bar{c}$, the steady state condition (8) determines the steady state level of the multiplier $\bar{\lambda}$.

It remains to calibrate the parameters ν and η. To that end, write condition (9) as

$$\eta\bar{l}^{1+\frac{1}{\nu}}=(1-\alpha)\,\frac{\bar{y}}{\bar{c}}.$$

In this equation we know $\bar{l}$, $\bar{c}$, and $\bar{y}$. We thus have one equation for the two parameters η and ν. The Frisch elasticity ν is typically calibrated based on microeconomic studies of labor supply elasticity. There is controversy regarding the value for ν. We will set $\nu=1$. Given this, we can use the previous equation to determine the constant η.

How do we calibrate the parameters of the stochastic process ρ and σ_{ε}^2? This may be done running a first order autoregression on estimated Solow residuals. Another possibility is to set ρ to some number and then choose σ_{ε}^2 to match the volatility of output in the data. For this latter approach we must solve the model many times for different values of σ_{ε}^2 and choose the one that makes the volatility of simulated output to match that observed in the data.

Log-linearization of the model

We will solve the model approximating the policy functions around the steady state. Most economists choose to log-linearize rather than to linearize their models. This gives log-linear equations that often seem to better describe data. Furthermore, log-linear policy functions have a nice economic interpretation as the percentage deviation of the particular variable being considered from its steady state value. Therefore, define for any (positive) variable x_t, its log-deviation from the steady state value

$$\hat{x}_t = \log\left(x_t/\bar{x}\right).$$

This implies that the level of the variable can be written in terms of the its log-deviation as

$$x_t = \bar{x}e^{\hat{x}_t}.$$

We will linearize the equilibrium conditions around $\hat{x}_t = 0$ for all varibles x_t.

Equation (1):
Write equation (1) in terms of the log-deviation from the steady state as

$$0 = \frac{1}{\bar{c}}e^{-\hat{c}_t} - \bar{\lambda}e^{\hat{\lambda}_t}$$

Performing a first order Taylor expansion around $(\hat{c}_t, \hat{\lambda}_t) = (0, 0)$ gives

$$0 \approx \frac{1}{\bar{c}} - \bar{\lambda} - \frac{1}{\bar{c}}\hat{c}_t - \bar{\lambda}\hat{\lambda}_t.$$

Using that in steady state $\frac{1}{\bar{c}} = \bar{\lambda}$ we obtain

$$0 \approx \hat{c}_t + \hat{\lambda}_t. \tag{17}$$

Note that the constant of the Taylor expansion dissapears because it is zero in steady state. This happens in all the equations and, therefore, from now on we ignore that constant term of the Taylor expansion.

Equation (2):
Write equation (2) in terms of the log-deviations from the steady state

$$0 = \eta\bar{l}^{\frac{1}{\nu}}e^{\frac{1}{\nu}\hat{l}_t} - \bar{\lambda}\left(1-\alpha\right)\frac{\bar{y}}{\bar{l}}e^{\hat{\lambda}_t+\hat{y}_t-\hat{l}_t}$$

Linearizing around $(\hat{l}_t, \hat{\lambda}_t, \hat{y}_t) = (0, 0, 0)$ gives

$$0 \approx \eta \bar{l}^{\frac{1}{\nu}} \frac{1}{\nu} \hat{l}_t - \bar{\lambda} (1 - \alpha) \frac{\bar{y}}{\bar{l}} [\hat{\lambda}_t + \hat{y}_t - \hat{l}_t]$$

But in steady state $\eta \bar{l}^{\frac{1}{\nu}} = \bar{\lambda} (1 - \alpha) \left(\bar{y}/\bar{l}\right)$, therefore

$$0 \approx \frac{1}{\nu} \hat{l}_t - \hat{\lambda}_t - \hat{y}_t + \hat{l}_t$$

or

$$0 \approx (1 + \frac{1}{\nu}) \hat{l}_t - \hat{\lambda}_t - \hat{y}_t. \tag{18}$$

Equation (3): Disregard for the moment the expectation operator—we will put it back later—and write the equation as

$$0 = \beta \bar{\lambda} e^{\hat{\lambda}_{t+1}} \left(\alpha \left(\bar{y}/\bar{k}\right) e^{\hat{y}_{t+1} - \hat{k}_{t+1}} + 1 - \delta \right) - \bar{\lambda} e^{\hat{\lambda}_t}.$$

Linearizing about $\left(\hat{\lambda}_{t+1}, \hat{y}_{t+1}, \hat{k}_{t+1}, \hat{\lambda}_t\right) = (0, 0, 0, 0)$ we obtain

$$0 \approx \beta \bar{\lambda} \left(\alpha \left(\bar{y}/\bar{k}\right) + 1 - \delta \right) \hat{\lambda}_{t+1} + \beta \bar{\lambda} \alpha \left(\bar{y}/\bar{k}\right) \left(\hat{y}_{t+1} - \hat{k}_{t+1} \right) - \bar{\lambda} \hat{\lambda}_t.$$

Dividing by $\bar{\lambda}$ and using that in steady state $\beta \left(\alpha \left(\bar{y}/\bar{k}\right) + 1 - \delta \right) = 1$ we have

$$0 \approx \hat{\lambda}_{t+1} + \beta \alpha \left(\bar{y}/\bar{k}\right) \left(\hat{y}_{t+1} - \hat{k}_{t+1} \right) - \hat{\lambda}_t.$$

Putting back the expectation operator we have

$$0 \approx E_t \left[\hat{\lambda}_{t+1} + \beta \alpha \left(\bar{y}/\bar{k}\right) \left(\hat{y}_{t+1} - \hat{k}_{t+1} \right) \right] - \hat{\lambda}_t \tag{19}$$

Equation (4). This equation is already log-linear. Taking the logarithm of the equation gives

$$\log y_t = \log A_t + \alpha \log k_t + (1 - \alpha) \log l_t$$

Subtracting the same equation at the steady state and rearranging gives

$$0 = \hat{y}_t - \hat{A}_t - \alpha \hat{k}_t - (1 - \alpha) \hat{l}_t \tag{20}$$

Equation (5):

$$0 = \bar{y}e^{\hat{y}_t} - \bar{c}e^{\hat{c}_t} - \bar{x}e^{\hat{x}_t}$$

Linearizing about $(\hat{y}_t, \hat{c}_t, \hat{x}_t) = (0, 0, 0)$ gives

$$0 \approx \bar{y}\hat{y}_t - \bar{c}\hat{c}_t - \bar{x}\hat{x}_t \tag{21}$$

Equation (6)

$$0 = \bar{k}e^{\hat{k}_{t+1}} - (1-\delta)\,\bar{k}e^{\hat{k}_t} - \bar{x}e^{\hat{x}_t}$$

Linearizing this equation gives

$$0 \approx \bar{k}\hat{k}_{t+1} - (1-\delta)\,\bar{k}\hat{k}_t - \bar{x}\hat{x}_t$$

But in steady state $\bar{x} = \delta\bar{k}$ which implies

$$0 \approx \hat{k}_{t+1} - (1-\delta)\,\hat{k}_t - \delta\hat{x}_t. \tag{22}$$

Equation (7)

Finally, the TFP equation is already loglinear and given by

$$0 = \log A_{t+1} - \rho \log A_t - \varepsilon_{t+1}.$$

Subtrating the same equation at the steady state and taking the conditional expectation as of time t on both side of the equation then gives

$$0 = E_t[\hat{A}_{t+1}] - \rho\hat{A}_t \tag{23}$$

Summarizing, the linearized rational expectations model can be written as follows (changing

somewhat the order of the equations)

$$
\begin{aligned}
0 &= \hat{c}_t + \hat{\lambda}_t && (24)\\
0 &= (1+\frac{1}{\nu})\hat{l}_t - \hat{\lambda}_t - \hat{y}_t && (25)\\
0 &= \hat{y}_t - \hat{A}_t - \alpha\hat{k}_t - (1-\alpha)\,\hat{l}_t && (26)\\
0 &= \bar{y}\hat{y}_t - \bar{c}\hat{c}_t - \bar{x}\hat{x}_t && (27)\\
E_t[\hat{k}_{t+1}] &= (1-\delta)\,\hat{k}_t + \delta\hat{x}_t && (28)\\
E_t\left[\hat{\lambda}_{t+1} + \beta\alpha\left(\bar{y}/\bar{k}\right)\left(\hat{y}_{t+1} - \hat{k}_{t+1}\right)\right] &= \lambda_t && (29)\\
E_t[\hat{A}_{t+1}] &= \rho\hat{A}_t && (30)
\end{aligned}
$$

Note that I wrode $E_t\left[\hat{k}_{t+1}\right]$ even though $\hat{k}_{t+1}$ is chosen (and therefore aldeary known) at time t. This is just notation that will allow us to write the model as the following first order vector expectational difference equation

$$\mathbf{A}E_t\left[\mathbf{z}_{t+1}\right] = \mathbf{B}\mathbf{z}_t \tag{31}$$

where the vector $\mathbf{z}_t$ contains all the variables in the economy and $\mathbf{A}$ and $\mathbf{B}$ are square matrices.

We solve numerically this model using the Matlab program `solab.m`. We order the variables $\mathbf{z}_t$ as follows:

$$\mathbf{z}_t = \begin{bmatrix} \text{endogenous states variables} \\ \text{exogenous states variables} \\ \text{jump variables} \end{bmatrix}$$

In the RBC model described above, the only endogenous state variable is the stock of capital $\hat{k}_t$ and the only exogenous state variable is the level of technology $\hat{A}_t$. Therefore, the variable $\mathbf{z}_t$ is given by

$$\mathbf{z}_t = \left[\hat{k}_t,\ \hat{A}_t,\ \ \hat{y}_t,\ \hat{c}_t,\ \hat{l}_t,\ \hat{x}_t,\ \hat{\lambda}_t\right]'. \tag{32}$$

Please note that the order within each group of variables does not matter (e.g. we could put c_t before y_t in the vector $\mathbf{z}_t$).

In addition, we must tell the program how many of the variables in $\mathbf{z}_t$ are state variables. This is to check the Blanchard-Kahn conditions to determine if the model has a unique solution or not. In our case, it is 2: $\hat{k}_t$ and $\hat{A}_t$. Note that, in this case, $\mathbf{A}$ and $\mathbf{B}$ are 7×7 matrices. If we let $\boldsymbol{\kappa}_t \equiv [\hat{k}_t, \hat{A}_t]'$ denote the vector of state variables and $\mathbf{u}_t = [\hat{y}_t,\ \hat{c}_t,\ \hat{l}_t,\ \hat{x}_t,\ \hat{\lambda}_t]$, the vector of jump variables, the solver delivers the equilibrium of the "certainty equivalent" model in the

form

$$\begin{aligned} \mathbf{u}_t &= \mathbf{F}\boldsymbol{\kappa}_t \\ \boldsymbol{\kappa}_{t+1} &= \mathbf{P}\boldsymbol{\kappa}_t \end{aligned}$$

The "stochastic" solution of the model is obtained by replacing the second equation above with

$$\boldsymbol{\kappa}_{t+1} = \mathbf{P}\boldsymbol{\kappa}_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \varepsilon_{t+1}$$

which simply recovers the stochastic shock $\hat{A}_{t+1} = \rho\hat{A}_t + \varepsilon_{t+1}$.

For the ordering (32), the matrices $\mathbf{A}$ and $\mathbf{B}$ of the system (31) are given by

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\beta\alpha\left(\bar{y}/\bar{k}\right) & 0 & \beta\alpha\left(\bar{y}/\bar{k}\right) & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \text{equation (24)} \\ \text{equation (25)} \\ \text{equation (26)} \\ \text{equation (27)} \\ \text{equation (28)} \\ \text{equation (29)} \\ \text{equation (30)} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & (1+\frac{1}{\nu}) & 0 & -1 \\ -\alpha & -1 & 1 & 0 & -(1-\alpha) & 0 & 0 \\ 0 & 0 & \bar{y} & -\bar{c} & 0 & -\bar{x} & 0 \\ 1-\delta & 0 & 0 & 0 & 0 & \delta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \rho & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \text{equation (24)} \\ \text{equation (25)} \\ \text{equation (26)} \\ \text{equation (27)} \\ \text{equation (28)} \\ \text{equation (29)} \\ \text{equation (30)} \end{bmatrix}$$

Using the calibrated parameter values, the model delivers the following solution:

$$\mathbf{F} = \begin{bmatrix} 0.22 & 1.33 \\ 0.57 & 0.34 \\ -0.17 & 0.50 \\ -1.10 & 5.07 \\ -0.57 & -0.34 \end{bmatrix}; \quad \mathbf{P} = \begin{bmatrix} 0.96 & 0.09 \\ 0 & 0.95 \end{bmatrix}$$

Which, in other words, implies the following policy functions:

$$\begin{aligned}
\hat{y}_t &= 0.22\hat{k}_t + 1.33\hat{A}_t \\
\hat{c}_t &= 0.57\hat{k}_t + 0.34\hat{A}_t \\
\hat{l}_t &= -0.17\hat{k}_t + 0.50\hat{A}_t \\
\hat{x}_t &= -1.10\hat{k}_t + 5.07\hat{A}_t \\
\hat{k}_{t+1} &= 0.96\hat{k}_t + 0.09\hat{A}_t \\
\hat{A}_{t+1} &= 0.95\hat{A}_t + \varepsilon_{t+1}.
\end{aligned}$$

Once we have this solution, we can compute impulse responses, variance decompositions, simulations, compute spectral densities, and so forth.

SIMPLE NEW KEYNESIAN MODEL

This note derives a simple New Keynesian (NK) model that produces a modern version of the classical Keynesian model with an IS curve, a Phillips curve, and a monetary policy rule. The model is a standard RBC model—in its simplest form, without capital—augmented with price frictions. The households' side of the model is identical to that in standard models. The difference is in the production side of the economy.

To model price stickiness we need to depart from the assumption of perfect competition. Indeed, price stickiness requires firms to set prices while perfect competition imposes price taking behavior. New Keynesian models assume that at least one sector is composed of monopolistic competitive firms with some pricing power. Furthermore, NK models assume that, for some reason, changing nominal prices is costly, or that some firms are simply unable to reset their prices every single period.

The model consists of a representatitive household, a final goods producing firm, a continuum of intermediate good producers (this is the sector with monopolistic power and price stickiness), and a monetary authority. In the end we obtain a very simple system of three linear difference equations. But the process of deriving them requires some work.

Households

The household sector is standard. We consider a model with money in the utility function in order to motivate a demand for currency (some authors even ignore the demand for money). The utility function of the representative household is

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\sigma}}{1-\sigma} - \eta \frac{l_t^{1+\nu}}{1+\nu} + \psi \frac{m_t^{1-\xi}}{1-\xi} \right)$$

where c_t is consumption, l_t is labor, $m_t = M_t/P_t$ denotes the real money demand, and β is the discount factor. The parameters σ, ν, η, ξ are all positive.

The budget constraint of the household in nominal terms is

$$P_t c_t + B_t + M_t = W_t l_t + D_t + R_{t-1} B_{t-1} + M_{t-1}$$

where P_t is the nominal price of goods, B_t are nominal bonds, M_t is the nominal demand for money, W_t is the nominal wage, R_{t-1} is the nominal gross interest rate on bonds between periods $t-1$ and t, and D_t are the profits from the firms owned by the household.

It is convenient to write the constraint in real terms. Dividing both sides of the previous

equation by P_t gives

$$c_t + b_t + m_t = w_t l_t + d_t + \frac{R_{t-1}}{\pi_t} b_{t-1} + \frac{m_{t-1}}{\pi_t}.$$

where $w_t = W_t/P_t$ is the real wage rate, $b_t = B_t/P_t$ is the demand for bonds in real terms, d_t are the real dividends, and π_t denotes the **gross** inflation rate between periods $t-1$ and t (that is $\pi_t = P_t/P_{t-1}$).

The problem of the consumer is to maximize his utility subject to a sequence of budget constraint. Using $\beta^t \lambda_t$ for the Lagrange multiplier on the budget constraint, the Lagrangian for the household's problem can be written as

$$L = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{array}{c} \frac{c_t^{1-\sigma}}{1-\sigma} - \eta \frac{l_t^{1+\nu}}{1+\nu} + \psi \frac{m_t^{1-\xi}}{1-\xi} \\ -\lambda_t \left[c_t + b_t + m_t - w_t l_t - d_t - R_{t-1} \frac{b_{t-1}}{\pi_t} - \frac{m_{t-1}}{\pi_t} \right] \end{array} \right\}.$$

The first order conditions are

$$\frac{\partial L}{\partial c_t} = 0 \Leftrightarrow c_t^{-\sigma} = \lambda_t$$

$$\frac{\partial L}{\partial l_t} = 0 \Leftrightarrow \eta l_t^{\nu} = \lambda_t w_t$$

$$\frac{\partial L}{\partial b_t} = 0 \Leftrightarrow \lambda_t = \beta E_t \left[\lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right]$$

$$\frac{\partial L}{\partial m_t} = 0 \Leftrightarrow \psi m_t^{-\xi} = \lambda_t - \beta E_t \left[\frac{\lambda_{t+1}}{\pi_{t+1}} \right]$$

However, because R_t is known at time t, the first order condition for b_t can be written as

$$\frac{\lambda_t}{R_t} = \beta E_t \left[\frac{\lambda_{t+1}}{\pi_{t+1}} \right]$$

Replacing this expression into the first order condition with respect to m_t gives

$$\psi m_t^{-\xi} = \lambda_t \frac{R_t - 1}{R_t}.$$

We can eliminate the multiplier λ_t and write the first order conditions as

$$\eta l_t^{\nu} = c_t^{-\sigma} w_t \tag{1}$$

$$\psi m_t^{-\xi} = c_t^{-\sigma} \frac{R_t - 1}{R_t} \tag{2}$$

$$c_t^{-\sigma} = \beta E_t \left[c_{t+1}^{-\sigma} \frac{R_t}{\pi_{t+1}} \right] \tag{3}$$

Production sector

There are two producing sectors: a competitive sector that produces final goods and an intermediate goods sector with a large number (a continuum) of firms each of which produces a differentiated intermediate input.

Final good firms.—

There is a perfectly competitive representative firm that produces final goods using a continuum of intermediate goods indexed by numbers in the unit interval $[0,1]$. The production function is given by

$$y_t = \left(\int_0^1 y_t(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}}, \tag{4}$$

where $\theta > 1$ represents the elasticity of substitution between pairs of intermediate inputs. Note that this production function has constant returns to scale and, therefore, firms earn zero profits in equilibrium.

The firm takes all prices as given, in particular, the nominal input prices $P_t(j)$ as well as the output price P_t. The problem of the firm is to maximize profits period by period. The objective function in nominal terms is

$$\max_{y_t(j)} P_t y_t - \int_0^1 P_t(j) y_t(j) dj.$$

That is, the firm maximizes total revenues, $P_t y_t$, minus the total cost of the inputs that it uses, $\int_0^1 P_t(j) y_t(j) dj$. Introducing the production function (4) into the objective function gives

$$\max_{y_t(j)} P_t \left(\int_0^1 y_t(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} - \int_0^1 P_t(j) y_t(j) dj.$$

The first order condition of this problem is to take the derivative with respect to $y_t(j)$ for all j and set it to zero. That is,

$$P_t \frac{\theta}{\theta-1} \left(\int_0^1 y_t(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}-1} \frac{\theta-1}{\theta} y_t(j)^{\frac{-1}{\theta}} - P_t(j) = 0 \text{ for all } j.$$

But note that

$$\left(\int_0^1 y_t(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}-1} = \left(\int_0^1 y_t(j)^{\frac{\theta-1}{\theta}} dj \right)^{\frac{1}{\theta-1}} = y_t^{\frac{1}{\theta}}.$$

Therefore, we can write the first order condition as

$$P_t y_t^{\frac{1}{\theta}} y_t(j)^{\frac{-1}{\theta}} = P_t(j).$$

From here we obtain the demands for the input $y_t(j)$ conditional on the level of output y_t

$$y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-\theta} y_t. \tag{5}$$

This equation says that the demand for each intermediate good $j \in [0,1]$ depends negatively on the relative price of the input and positively on the level of production. Note that θ is also the demand elasticity of the intermediate good j.

The firm has constant returns to scale, so its profits must be zero. To see this formally, note that profits satisfy

$$\begin{aligned}
\text{Profits}_t &= P_t y_t - \int_0^1 P_t(j)\, y_t(j)\, dj \\
&= P_t y_t - \int_0^1 P_t(j) \left(\frac{P_t(j)}{P_t}\right)^{-\theta} y_t dj \\
&= \left[P_t - P_t^{\theta} \int_0^1 P_t(j)^{1-\theta} dj\right] y_t.
\end{aligned}$$

The term in square brackets does not depend on the firm's choices because the firm is a price taker. Suppose that the term in square brackets is negative. Then any level of positive output $y_t > 0$ will generate negative profits. Therefore, it will be optimal for the firm to choose $y_t = 0$. But this can't be an equilibrium because consumption will be zero and the marginal utility of consumption at zero is infinity. Therefore, the term in square brackets cannot be negative. Suppose now that the term in square brackets is positive. Then the firm could make infinite profits by producing an infinite amount of goods. This also can't be an equilibrium because this will violate feasibility (production is limited by the endowment of labor). Therefore, the only possibility is that the term in square brackets is zero and the firm makes zero profits. Equating the term in square brackets to zero gives

$$P_t = P_t^{\theta} \int_0^1 P_t(j)^{1-\theta} dj$$

or

$$P_t = \left[\int_0^1 P_t(j)^{1-\theta} dj\right]^{\frac{1}{1-\theta}} \tag{6}$$

This expression can be interpreted as a price index of the final good.

Intermediate good firms.—

There is a continuum of intermediate good firms indexed by $j \in [0,1]$ each of which is a price setter. Firm j produces an intermediate good using labor as the only input of production. The production function is given by

$$y_t(j) = A_t l_t^d(j) \tag{7}$$

where A_t is the level of technology, common across of intermediate good firms, and $l_t^d(j)$ is the amount of labor used by firm j. Labor markets are perfectly competitive.

Each firm is monopolistic in that it chooses its optimal price $P_t(j)$ internalizing that the demand they face is given by (5). To introduce price stickiness, we assume that firms face nominal rigidities in terms of a quadratic price adjustment cost

$$AC_t(j) = \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2 y_t. \tag{8}$$

The parameter ω governs the price stickiness in the economy and $\bar{\pi}$ is the gross inflation rate in steady state. Note that the adjustment cost is measured in terms of the final consumption good. The assumption here is that if the firm chooses to increase or decrease the price it charges at a different rate from the long-run gross inflation rate $\bar{\pi}$, it incurrs in a quadratic adjustment cost.

The adjustment cost makes the firm's problem dynamic. The problem of the firm at time $t = 0$ is to maximize the present discounted value of future nominal dividends,

$$E_0 \sum_{t=0}^{\infty} Q_t D_t(j)$$

where Q_t is the nominal discount rate used by the firm to discount future nominal profits (more on this below) and

$$D_t(j) = P_t(j)\, y_t(j) - W_t l_t^d(j) - P_t \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2 y_t$$

is the nominal dividend in period t.

It is convenient to write the firm's problem in real terms. The real present value of all future nominal dividends is given by

$$\max_{\{l_t(j), y_t(j), P_t(j)\}} E_0 \sum_{t=0}^{\infty} \frac{Q_t}{P_0} D_t(j) = \max_{\{l_t(j), y_t(j), P_t(j)\}} E_0 \sum_{t=0}^{\infty} Q_t \frac{P_t}{P_0} d_t(j)$$

where

$$d_t = \frac{D_t}{P_t} = \frac{P_t(j)}{P_t} y_t(j) - w_t l_t^d(j) - \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2 y_t.$$

are the real dividends at time t and w_t is the real wage.

Note that $q_t \equiv Q_t \frac{P_t}{P_0}$ is the real discount rate that the firm uses to discount future real dividends. The discount rate measures the value that the firm attaches to a unit of real profits in period t—goods at time t—in terms of goods at time $t = 0$. Because the firm is owned by the household, the firm should discount future profits using the discount factor of its owners: the consumers. The discount factor of the consumers is the marginal rate of substitution between goods at time t in terms of goods at time 0. Therefore, the relevant real discount rate for the firm is

$$q_t = \beta^t \frac{c_t^{-\sigma}}{c_0^{-\sigma}}. \tag{9}$$

Using this insight, the firm's problem can be written as

$$\max_{P_t(j), y_t(j), l_t(j)} E_0 \sum_{t=0}^{\infty} q_t \left[\frac{P_t(j)}{P_t} y_t(j) - w_t l_t^d(j) - \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2 y_t\right]$$

subject to the demand (5), the technology (7), and the adjustment cost (8).

We first use (7) and solve for labor as a function of output

$$l_t^d(j) = \frac{y_t(j)}{A_t}.$$

Therefore, real profits of firm j at time t are

$$d_t(j) = \frac{P_t(j)}{P_t} y_t(j) - \frac{w_t}{A_t} y_t(j) - \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2 y_t$$

Note that the term w_t/A_t is the real marginal cost of production.

Next, we use the demand function (5) to replace $y_t(j)$ into the previous expression and write profits as

$$\begin{aligned} d_t(j) &= \frac{P_t(j)}{P_t}\left(\frac{P_t(j)}{P_t}\right)^{-\theta} y_t - \frac{w_t}{A_t}\left(\frac{P_t(j)}{P_t}\right)^{-\theta} y_t - \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2 y_t \\ &= y_t\left[\frac{P_t(j)^{1-\theta}}{P_t^{1-\theta}} - \frac{w_t}{A_t}\frac{P_t(j)^{-\theta}}{P_t^{-\theta}} - \frac{\omega}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi}\right)^2\right] \end{aligned}$$

It then follows that the present discounted value of future real profits is

$$\max_{P_t(j)} E_0 \sum_{t=0}^{\infty} q_t y_t \left[\frac{P_t(j)^{1-\theta}}{P_t^{1-\theta}} - \frac{w_t}{A_t} \frac{P_t(j)^{-\theta}}{P_t^{-\theta}} - \frac{\omega}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi} \right)^2 \right].$$

The maximization is taken with respecto to the nominal price $P_t(j)$.

The first order condition with respect to $P_t(j)$ is

$$\begin{aligned} 0 &= q_t y_t \left[(1-\theta) \frac{P_t(j)^{-\theta}}{P_t^{1-\theta}} + \theta \frac{w_t}{A_t} \frac{P_t(j)^{-\theta-1}}{P_t^{-\theta}} - \omega \left(\frac{P_t(j)}{P_{t-1}(j)} - \bar{\pi} \right) \frac{1}{P_{t-1}(j)} \right] \\ &\quad + E_t \left[q_{t+1} y_{t+1} \omega \left(\frac{P_{t+1}(j)}{P_t(j)} - \bar{\pi} \right) \frac{P_{t+1}(j)}{P_t(j)^2} \right]. \end{aligned} \tag{10}$$

Note that if $\omega = 0$—so that there are no adjustment costs—the above expression reduces to

$$0 = q_t y_t \left[(1-\theta) \frac{P_t(j)^{-\theta}}{P_t^{1-\theta}} + \theta \frac{w_t}{A_t} \frac{P_t(j)^{-\theta-1}}{P_t^{-\theta}} \right]$$

or

$$P_t(j) = \frac{\theta}{\theta - 1} \frac{P_t w_t}{A_t} = \frac{\theta}{\theta - 1} \frac{W_t}{A_t}$$

which says that the price is set as a constant mark-up over the nominal marginal cost. This is the usual result with a monopolistic producer facing a demand with a constant elasticity of θ.

Monetary Authority

We assume that monetary policy is described by a (modified) Taylor rule of the form

$$\log(R_t) = (1-\rho_R) \log R_t^* + \rho_R \log R_{t-1} + v_t. \tag{11}$$

where $\log R_t$ is the net nominal interest rate (we always use $\log(\cdot)$ for the natural logarithm), $\log R_t^*$ is the nominal target rate, ρ_R is a smoothing parameter, and v_t is a monetary policy shock to be described below.

We assume that the target interest rate satisfies

$$\log R_t^* = \log \bar{R} + \phi_\pi \log \left(\frac{\pi_t}{\bar{\pi}} \right) + \phi_y \log \left(\frac{y_t}{\bar{y}} \right) \tag{12}$$

where $\bar{R}$, $\bar{\pi}$, and $\bar{y}$ are the steady state values of the gross nominal interest rate, inflation, and output respectively. In this specification, the target rate reacts to deviations of inflation and

output from their steady state values $\bar{\pi}$ and $\bar{y}$. The higher is ϕ_π, the stronter is the response of the target rate to deviations of inflation from its long run value. The same interpretation holds for ϕ_y.

Plugging (12) into (11) gives the following Taylor rule

$$\log(R_t) = (1-\rho_R)\left[\log\bar{R} + \phi_\pi \log\left(\frac{\pi_t}{\bar{\pi}}\right) + \phi_y \log\left(\frac{y_t}{\bar{y}}\right)\right] + \rho_R \log R_{t-1} + v_t. \tag{13}$$

Shocks

We assume that technology and monetary policy shocks evolve according to first order autoregressive processes

$$\log A_{t+1} = \rho_A \log A_t + \varepsilon_{A,t+1}, \tag{14}$$

$$v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1}. \tag{15}$$

where $\varepsilon_{A,t+1}$ and $\varepsilon_{v,t+1}$ are uncorrelated i.i.d. shocks with a standard deviation σ_A and σ_v. The innovations $\varepsilon_{A,t}$ and ε_{vt} are uncorrelated at all leads and lags.

Equilibrium

We focus on a **symmetric equilibrium** in which all intermediate good producers make identical choices. This implies that the subscript j in the corresponding expressions can be dropped. Moreover, expression (6) implies that the aggregate price level P_t is equal to the price chosen by the intermediate good producers, and condition (5) implies $y_t(j) = y_t$ for all j.

Feasibility in goods and labor markets are, respectively

$$y_t = c_t + AC_t \text{ and } l_t = l_t^d \tag{16}$$

Note that feasibility in goods markets include the adjustment cost since it is incurred in terms of final consumption goods.

We now perform some algebra on the price setting condition (10). Using **symmetry** implies that $P_t(j) = P_t$ for all j,

$$\begin{aligned} 0 &= q_t y_t \left[(1-\theta)\frac{P_t^{-\theta}}{P_t^{1-\theta}} + \theta\frac{w_t}{A_t}\frac{P_t^{-\theta-1}}{P_t^{-\theta}} - \omega\left(\frac{P_t}{P_{t-1}} - \bar{\pi}\right)\frac{1}{P_{t-1}}\right] \\ &\quad + E_t\left[q_{t+1} y_{t+1}\omega\left(\frac{P_{t+1}}{P_t} - \bar{\pi}\right)\frac{P_{t+1}}{P_t^2}\right]. \end{aligned}$$

Canceling terms,

$$\begin{aligned} 0 &= q_t y_t \left[(1-\theta) \frac{1}{P_t} + \theta \frac{w_t}{A_t} \frac{1}{P_t} - \omega \left(\pi_t - \bar{\pi} \right) \frac{1}{P_{t-1}} \right] \\ &\quad + E_t \left[q_{t+1} y_{t+1} \omega \left(\pi_{t+1} - \bar{\pi} \right) \pi_{t+1} \frac{1}{P_t} \right]. \end{aligned}$$

Multiplying the previous expression by P_t gives

$$0 = q_t y_t \left[(1-\theta) + \theta \frac{w_t}{A_t} - \omega \left(\pi_t - \bar{\pi} \right) \pi_t \right] + E_t \left[q_{t+1} y_{t+1} \omega \left(\pi_{t+1} - \bar{\pi} \right) \pi_{t+1} \right].$$

Using $q_t = \beta^t c_t^{-\sigma} / c_0^{-\sigma}$ gives

$$0 = \beta^t \frac{c_t^{-\sigma}}{c_0^{-\sigma}} y_t \left[(1-\theta) + \theta \frac{w_t}{A_t} - \omega \left(\pi_t - \bar{\pi} \right) \pi_t \right] + E_t \left[\beta^{t+1} \frac{c_{t+1}^{-\sigma}}{c_0^{-\sigma}} y_{t+1} \omega \left(\pi_{t+1} - \bar{\pi} \right) \pi_{t+1} \right].$$

Dividing both sides of the equation by $\beta^t \left(c_t^{-\sigma} / c_0^{-\sigma} \right) y_t$ and rearranging gives,

$$\left(\pi_t - \bar{\pi} \right) \pi_t = \frac{\theta}{\omega} \left[\frac{w_t}{A_t} - \frac{\theta - 1}{\theta} \right] + E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\sigma} \frac{y_{t+1}}{y_t} \left(\pi_{t+1} - \bar{\pi} \right) \pi_{t+1} \right] \tag{17}$$

Therefore the equilibrium conditions of the model are summarized by the following system of nine equations

$$\eta l_t^{\nu} = c_t^{-\sigma} w_t \tag{18}$$

$$\psi m_t^{-\xi} = c_t^{-\sigma} \frac{R_t - 1}{R_t} \tag{19}$$

$$c_t^{-\sigma} = \beta E_t \left[c_{t+1}^{-\sigma} \frac{R_t}{\pi_{t+1}} \right] \tag{20}$$

$$y_t = A_t l_t \tag{21}$$

$$y_t = c_t + \frac{\omega}{2} \left(\pi_t - \bar{\pi} \right)^2 y_t \tag{22}$$

$$\left(\pi_t - \bar{\pi} \right) \pi_t = \frac{\theta}{\omega} \left[\frac{w_t}{A_t} - \frac{\theta - 1}{\theta} \right] + E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\sigma} \frac{y_{t+1}}{y_t} \left(\pi_{t+1} - \bar{\pi} \right) \pi_{t+1} \right] \tag{23}$$

$$\log \left(R_t \right) = \left(1 - \rho_R \right) \left[\log \bar{R} + \phi_\pi \log \left(\frac{\pi_t}{\bar{\pi}} \right) + \phi_y \log \left(\frac{y_t}{\bar{y}} \right) \right] + \rho_R \log R_{t-1} + v_t \tag{24}$$

$$\log A_{t+1} = \rho_A \log A_t + \varepsilon_{A,t+1} \tag{25}$$

$$v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1} \tag{26}$$

We want to simplify this system. Use (18) and (21) to find the real marginal cost w_t/A_t as a function of output, consumption, and the level of technology

$$\frac{w_t}{A_t} \equiv mc_t = \frac{\eta l_t^{\nu} c_t^{\sigma}}{A_t} = \eta \frac{y_t^{\nu} c_t^{\sigma}}{A_t^{1+\nu}}. \tag{27}$$

Inserting this equation into the pricing condition (23) gives

$$(\pi_t - \bar{\pi})\,\pi_t = \frac{\theta}{\omega}\left[\eta \frac{y_t^{\nu} c_t^{\sigma}}{A_t^{1+\nu}} - \frac{\theta - 1}{\theta}\right] + E_t\left[\beta \left(\frac{c_{t+1}}{c_t}\right)^{-\sigma} \frac{y_{t+1}}{y_t}\,(\pi_{t+1} - \bar{\pi})\,\pi_{t+1}\right]$$

Therefore, we have reduced the system to the following seven equations

$$\psi m_t^{-\xi} = c_t^{-\sigma} \frac{R_t - 1}{R_t} \tag{28}$$

$$c_t^{-\sigma} = \beta E_t\left[c_{t+1}^{-\sigma} \frac{R_t}{\pi_{t+1}}\right] \tag{29}$$

$$y_t\left[1 - \frac{\omega}{2}\,(\pi_t - \bar{\pi})^2\right] = c_t \tag{30}$$

$$(\pi_t - \bar{\pi})\,\pi_t = \frac{\theta}{\omega}\left[\eta \frac{y_t^{\nu} c_t^{\sigma}}{A_t^{1+\nu}} - \frac{\theta - 1}{\theta}\right] + E_t\left[\beta \left(\frac{c_{t+1}}{c_t}\right)^{-\sigma} \frac{y_{t+1}}{y_t}\,(\pi_{t+1} - \bar{\pi})\,\pi_{t+1}\right] \tag{31}$$

$$\log R_t = (1 - \rho_R)\left[\log \bar{R} + \phi_{\pi} \log\left(\frac{\pi_t}{\bar{\pi}}\right) + \phi_y \log\left(\frac{y_t}{\bar{y}}\right)\right] + \rho_R \log R_{t-1} + v_t \tag{32}$$

$$\log A_{t+1} = \rho_A \log A_t + \varepsilon_{A,t+1} \tag{33}$$

$$v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1} \tag{34}$$

This system can be reduced further, but it is convenient to leave it as it is and reduce it after we perform the linearization of the equilibrium conditions. To linearize the model, we first need to find the steady state

Steady state

Evaluating the previous system of equations at the steady state gives

$$\psi \bar{m}^{-\xi} = \bar{c}^{-\sigma} \frac{\bar{R} - 1}{\bar{R}}$$

$$\bar{c}^{-\sigma} = \beta \bar{c}^{-\sigma} \frac{\bar{R}}{\bar{\pi}}$$

$$\bar{y} = \bar{c}$$

$$(\bar{\pi} - \bar{\pi})\,\bar{\pi} = \frac{\theta}{\omega}\left[\eta \frac{\bar{y}^{\nu}\bar{c}^{\sigma}}{\bar{A}^{1+\nu}} - \frac{\theta - 1}{\theta}\right] + \beta \left(\frac{\bar{c}}{\bar{c}}\right)^{-\sigma} \frac{\bar{y}}{\bar{y}} (\bar{\pi} - \bar{\pi})\,\bar{\pi}$$

$$\log \bar{R} = (1 - \rho_R)\left[\log \bar{R} + \phi_{\pi} \log\left(\frac{\bar{\pi}}{\bar{\pi}}\right) + \phi_y \log\left(\frac{\bar{y}}{\bar{y}}\right)\right] + \rho_R \log \bar{R}$$

$$\log \bar{A} = \rho_A \log \bar{A}$$

$$\bar{v} = \rho_v \bar{v}$$

This can be solved for the steady state quantities and prices

$$\bar{R} = \bar{\pi}/\beta \tag{35}$$

$$\bar{A} = 1 \tag{36}$$

$$\bar{v} = 0 \tag{37}$$

$$\bar{c} = \bar{y} = \left(\frac{1}{\eta}\frac{\theta - 1}{\theta}\right)^{\frac{1}{\sigma+\nu}} \tag{38}$$

$$\bar{m} = \left[\psi \bar{y}^{\sigma} \frac{\bar{R}}{\bar{R} - 1}\right]^{1/\xi} \tag{39}$$

Note that the steady state interest rate and money demand depend on the steady state inflation rate, which is chosen by the monetary authority. For example, if the monetary authority targets zero inflation, so that $\bar{\pi} = 1$, then $\bar{R} = 1/\beta$.

Log-linearization of the equilibrium conditions

We log-linearize the equilibrium conditions around the steady state. To that end we define, for any variable x_t, its log-deviation from the steady state as

$$\hat{x}_t = \log\left(x_t/\bar{x}\right)$$

which implies

$$x_t = \bar{x} e^{\hat{x}_t}.$$

We now rewrite the system of equations (28)–(33) in terms of the transformed variables. We do not transform the monetary policy shock v_t, which is already linear with mean zero.

Equation (28):

$$\begin{aligned} 0 &= \bar{c}^{-\sigma} e^{-\sigma \hat{c}_t} \frac{\bar{R} e^{\hat{R}_t} - 1}{\bar{R} e^{\hat{R}_t}} - \psi \bar{m}^{-\xi} e^{-\xi \hat{m}_t} \\ &= \frac{\bar{c}^{-\sigma}}{\bar{R}} e^{-\sigma \hat{c}_t} \left(\bar{R} e^{\hat{R}_t} - 1 \right) e^{-\hat{R}_t} - \psi \bar{m}^{-\xi} e^{-\xi \hat{m}_t} \end{aligned}$$

Linearizing this equation around $\left(\hat{c}_t, \hat{R}_t, \hat{m}_t \right) = (0, 0, 0)$ (we ignore the constant because it is zero in steady state) gives

$$\begin{aligned} 0 &\approx \frac{\bar{c}^{-\sigma}}{\bar{R}} \left(-\sigma \hat{c}_t \right) \left(\bar{R} - 1 \right) + \frac{\bar{c}^{-\sigma}}{\bar{R}} \left[\bar{R} \hat{R}_t - \left(\bar{R} - 1 \right) \hat{R}_t \right] - \psi \bar{m}^{-\xi} \left(-\xi \hat{m}_t \right) \\ &\approx -\sigma \hat{c}_t \bar{c}^{-\sigma} \frac{\bar{R} - 1}{\bar{R}} + \bar{c}^{-\sigma} \frac{\bar{R} - 1}{\bar{R}} \frac{\hat{R}_t}{\bar{R} - 1} + \psi \bar{m}^{-\xi} \xi \hat{m}_t \end{aligned}$$

But in steady state

$$\bar{c}^{-\sigma} \frac{\bar{R} - 1}{\bar{R}} = \psi \bar{m}^{-\xi}$$

then we have

$$0 \approx -\sigma \hat{c}_t + \frac{\hat{R}_t}{\bar{R} - 1} + \xi \hat{m}_t \tag{40}$$

Equation (29):

$$\begin{aligned} 0 &= \beta E_t \left[c_{t+1}^{-\sigma} \frac{R_t}{\pi_{t+1}} \right] - c_t^{-\sigma} \\ &= \beta E_t \left[\bar{c}^{-\sigma} e^{-\sigma \hat{c}_{t+1}} \frac{\bar{R} e^{\hat{R}_t}}{\bar{\pi} e^{\hat{\pi}_{t+1}}} \right] - \bar{c}^{-\sigma} e^{-\sigma \hat{c}_t} \\ &= \beta E_t \left[\frac{\bar{c}^{-\sigma} \bar{R}}{\bar{\pi}} e^{-\sigma \hat{c}_{t+1} + \hat{R}_t - \hat{\pi}_{t+1}} \right] - \bar{c}^{-\sigma} e^{-\sigma \hat{c}_t} \end{aligned}$$

Ignoring for the moment the expectation operator (we put it back after doing the algebra) and linearizing around $\left(\hat{R}, \hat{c}_{t+1}, \hat{\pi}_{t+1}, \hat{c}_t \right) = (0, 0, 0, 0)$ gives

$$0 \approx \beta \frac{\bar{c}^{-\sigma} \bar{R}}{\bar{\pi}} \left[-\sigma \hat{c}_{t+1} + \hat{R}_t - \hat{\pi}_{t+1} \right] + \bar{c}^{-\sigma} \sigma \hat{c}_t$$

$$0 \approx \beta \bar{c}^{-\sigma} \frac{\bar{R}}{\bar{\pi}} \left[\hat{R}_t - \hat{\pi}_{t+1} - \sigma \hat{c}_{t+1} \right] + \bar{c}^{-\sigma} \sigma \hat{c}_t$$

But in steady state $\beta\bar{c}^{-\sigma}\frac{\bar{R}}{\bar{\pi}} = \bar{c}^{-\sigma}$ so that

$$0 \approx \hat{R}_t - \hat{\pi}_{t+1} - \sigma\hat{c}_{t+1} + \sigma\hat{c}_t$$

$$0 \approx \hat{R}_t - \hat{\pi}_{t+1} - \sigma\hat{c}_{t+1} + \sigma\hat{c}_t.$$

Reinserting the expectation operator and rearranging gives

$$E_t\hat{c}_{t+1} \approx \hat{c}_t + \frac{1}{\sigma}\left[\hat{R}_t - E_t\hat{\pi}_{t+1}\right]. \tag{41}$$

which is the linearized Euler equation.

Equation (30):

$$\begin{aligned} 0 &= y_t\left[1 - \frac{\omega}{2}\left(\pi_t - \bar{\pi}\right)^2\right] - c_t \\ &= \bar{y}e^{\hat{y}_t}\left[1 - \frac{\omega}{2}\left(\bar{\pi}e^{\hat{\pi}_t} - \bar{\pi}\right)^2\right] - \bar{c}e^{\hat{c}_t} \end{aligned}$$

Linearizing around $(\hat{y}_t, \hat{\pi}_t, \hat{c}_t) = (0, 0, 0)$ gives

$$0 \approx \bar{y}\hat{y}_t - \bar{y}\omega\left(\bar{\pi} - \bar{\pi}\right)\hat{\pi}_t - \bar{c}\hat{c}_t$$

Using $\bar{y} = \bar{c}$ we obtain

$$\hat{c}_t \approx \hat{y}_t. \tag{42}$$

Equation (31):

We can write this equation as

$$\begin{aligned} 0 &= \frac{\theta}{\omega}\left[\eta\frac{\bar{y}^\nu\bar{c}^\sigma}{\bar{A}^{1+\nu}}e^{\nu\hat{y}_t+\sigma\hat{c}_t-(1+\nu)\hat{A}_t} - \frac{\theta-1}{\theta}\right] - \left(\bar{\pi}e^{\hat{\pi}_t} - \bar{\pi}\right)\bar{\pi}e^{\hat{\pi}_t} \\ &\quad + E_t\left[\beta\bar{\pi}e^{-\sigma(\hat{c}_{t+1}-\hat{c}_t)+(\hat{y}_{t+1}-\hat{y}_t)+\hat{\pi}_{t+1}}\left(\bar{\pi}e^{\hat{\pi}_{t+1}} - \bar{\pi}\right)\right] \end{aligned}$$

Ignoring for the moment the expectation operator and linearizing around $\left(\hat{\pi}_t, \hat{y}_t, \hat{c}_t, \hat{A}_t, \hat{\pi}_{t+1}, \hat{y}_{t+1}, \hat{c}_{t+1}\right) = (0, 0, 0, 0, 0, 0, 0)$ gives

$$0 \approx \frac{\theta}{\omega}\eta\frac{\bar{y}^\nu\bar{c}^\sigma}{\bar{A}^{1+\nu}}\left[\nu\hat{y}_t + \sigma\hat{c}_t - (1+\nu)\hat{A}_t\right] - \bar{\pi}^2\hat{\pi}_t + \beta\bar{\pi}^2\hat{\pi}_{t+1}$$

but in steady state

$$\eta\frac{\bar{y}^\nu\bar{c}^\sigma}{\bar{A}^{1+\nu}} = \frac{\theta-1}{\theta}$$

Therefore we have

$$0 \approx \frac{\theta-1}{\omega}\left[\nu\hat{y}_t + \sigma\hat{c}_t - (1+\nu)\hat{A}_t\right] - \bar{\pi}^2\hat{\pi}_t + \beta\bar{\pi}^2\hat{\pi}_{t+1}$$

Dividing by $\bar{\pi}^2$, inserting the expectation operator, and rearranging

$$\beta E_t\left[\hat{\pi}_{t+1}\right] \approx \hat{\pi}_t - \frac{\theta-1}{\bar{\pi}^2\omega}\left[\nu\hat{y}_t + \sigma\hat{c}_t - (1+\nu)\hat{A}_t\right]$$

But using that $\hat{y}_t = \hat{c}_t$ (this is equation (42)), we can write this equation as

$$\hat{\pi}_t = \beta E_t\left[\hat{\pi}_{t+1}\right] + \kappa\left[(\nu+\sigma)\hat{y}_t - (1+\nu)\hat{A}_t\right] \tag{43}$$

where

$$\kappa \equiv \frac{\theta-1}{\bar{\pi}^2\omega}.$$

Please note that the term $\left[(\nu+\sigma)\hat{y}_t - (1+\nu)\hat{A}_t\right]$ is actually the log-deviation of the marginal cost of the firm from its steady state value

$$\widehat{mc}_t = (\nu+\sigma)\hat{y}_t - (1+\nu)\hat{A}_t.$$

Using this observation, you can write (43) as

$$\hat{\pi}_t = \beta E_t\hat{\pi}_{t+1} + \kappa\widehat{mc}_t \tag{44}$$

which is an expression that appears in many papers of the literature.

Equation (32):

This equation is already log-linear. Subtracting $\log\left(\bar{R}\right)$ from both sides we have

$$\log\left(\frac{R_t}{\bar{R}}\right) = (1-\rho_R)\left[\log\left(\frac{\bar{R}}{\bar{R}}\right) + \phi_\pi\log\left(\frac{\pi_t}{\bar{\pi}}\right) + \phi_y\log\left(\frac{y_t}{\bar{y}}\right)\right] + \rho_R\log\left(\frac{R_{t-1}}{\bar{R}}\right) + v_t$$

Using the definition of hatted variables,

$$\hat{R}_t = (1-\rho_R)\left[\phi_\pi\hat{\pi}_t + \phi_y\hat{y}_t\right] + \rho_R\hat{R}_{t-1} + v_t \tag{45}$$

Equation (33):

The equation for the evolution of productivity is already log-linear. We thus have,

$$\hat{A}_{t+1} = \rho_A\hat{A}_t + \varepsilon_{A,t+1} \tag{46}$$

We also keep the interest rate shock linear as in equation (34)

Summarizing, the linearized system of equations is

$$0 = -\sigma \hat{c}_t + \frac{\hat{R}_t}{\bar{R} - 1} + \xi \hat{m}_t$$

$$E_t \left[\hat{c}_{t+1}\right] = \hat{c}_t + \frac{1}{\sigma}\left[\hat{R}_t - E_t \left[\hat{\pi}_{t+1}\right]\right]$$

$$0 = \hat{c}_t - \hat{y}_t$$

$$\hat{\pi}_t = \beta E_t \left[\hat{\pi}_{t+1}\right] + \kappa \left[\left(\nu + \sigma\right)\hat{y}_t - \left(1 + \nu\right)\hat{A}_t\right]$$

$$\hat{R}_t = \left(1 - \rho_R\right)\left[\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right] + \rho_R \hat{R}_{t-1} + v_t$$

$$\hat{A}_{t+1} = \rho_A \hat{A}_t + \varepsilon_{A,t+1}$$

$$v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1}$$

This can be simplified. Use the third condition to get rid of all terms with $\hat{c}_t$:

$$0 = -\sigma \hat{y}_t + \frac{\hat{R}_t}{\bar{R} - 1} + \xi \hat{m}_t \tag{47}$$

$$E_t \left[\hat{y}_{t+1}\right] = \hat{y}_t + \frac{1}{\sigma}\left[\hat{R}_t - E_t \left[\hat{\pi}_{t+1}\right]\right] \tag{48}$$

$$\hat{\pi}_t = \beta E_t \left[\hat{\pi}_{t+1}\right] + \kappa \left[\left(\nu + \sigma\right)\hat{y}_t - \left(1 + \nu\right)\hat{A}_t\right] \tag{49}$$

$$\hat{R}_t = \left(1 - \rho_R\right)\left[\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right] + \rho_R \hat{R}_{t-1} + v_t \tag{50}$$

$$\hat{A}_{t+1} = \rho_A \hat{A}_t + \varepsilon_{A,t+1} \tag{51}$$

$$v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1} \tag{52}$$

The set of equations (48), (49), and (50) is usually referred to as *The Three Equation New Keynesian Model*. Given the equilibrium values of $\hat{y}_t$ and $\hat{R}_t$, equation (47) can be used to recover the money demand $\hat{m}_t$, so we basically can ignore it for the purpose of finding the equilibrium. Therefore, from now on we focus on the equations (48), (49), (50), (51), and (52).

We want to put this system of equations in a form to be used in Paul Klein's Matlab function `solab.m`. We need to identify the state and the control variables. The state or predetermined variables are those variables that agents cannot affect at time t and that help predict the future evolution of the system. In the current model, at the beginning of period t neither R_{t-1}, A_t, nor v_t can be changed by anyone. Furthermore, those variables help predict the future evolution of

the economy. Therefore, the state variables of this model are

$$\mathbf{x}_t = \left[\hat{R}_{t-1}, \hat{A}_t, v_t\right]'$$

The control variables are the rest of the variables in the economy—recall that we will recover the demand for money later on,

$$\mathbf{y}_t = [\hat{y}_t, \hat{\pi}_t]$$

so that

$$\mathbf{s}_{t+1} = \begin{bmatrix} \hat{R}_t \\ \hat{A}_{t+1} \\ v_{t+1} \\ \hat{y}_{t+1} \\ \hat{\pi}_{t+1} \end{bmatrix} \qquad \mathbf{s}_t = \begin{bmatrix} \hat{R}_{t-1} \\ \hat{A}_t \\ v_t \\ \hat{y}_t \\ \hat{\pi}_t \end{bmatrix}$$

We rewrite our system of equations in a form to be used in the `solab.m` routine

$$\begin{aligned} E_t\left[\hat{y}_{t+1}\right] + \frac{1}{\sigma} E_t\left[\hat{\pi}_{t+1}\right] - \frac{1}{\sigma} E_t\left[\hat{R}_t\right] &= \hat{y}_t \\ \beta E_t\left[\hat{\pi}_{t+1}\right] &= \hat{\pi}_t - \kappa\left[(\nu+\sigma)\,\hat{y}_t - (1+\nu)\,\hat{A}_t\right] \\ E_t\left[\hat{R}_t\right] &= (1-\rho_R)\left[\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right] + \rho_R \hat{R}_{t-1} + v_t \\ E_t\left[\hat{A}_{t+1}\right] &= \rho_A \hat{A}_t \\ E_t\left[v_{t+1}\right] &= \rho_v v_t. \end{aligned}$$

Note that we are adding the irrelevant expectation in the interest rate at time t ($R_t = E_t[R_t]$ is known at t,) because this is the notation that is used in the `solab.m` routine.

We want to put this system of equations in the form

$$\mathbf{A} E_t\left[\mathbf{s}_{t+1}\right] = \mathbf{B}\mathbf{s}_t$$

where $\mathbf{A}$ and $\mathbf{B}$ are 5×5 matrices. We have

$$\begin{bmatrix} -\frac{1}{\sigma} & 0 & 0 & 1 & \frac{1}{\sigma} \\ 0 & 0 & 0 & 0 & \beta \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} E_t \begin{bmatrix} \hat{R}_t \\ \hat{A}_{t+1} \\ v_{t+1} \\ \hat{y}_{t+1} \\ \hat{\pi}_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & \kappa(1+\nu) & 0 & -\kappa(\nu+\sigma) & 1 \\ \rho_R & 0 & 1 & (1-\rho_R)\,\phi_y & (1-\rho_R)\,\phi_\pi \\ 0 & \rho_A & 0 & 0 & 0 \\ 0 & 0 & \rho_v & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{R}_{t-1} \\ \hat{A}_t \\ v_t \\ \hat{y}_t \\ \hat{\pi}_t \end{bmatrix}.$$

The Blanchard and Kahn condition requires that this system has exactly 3 stable generalized

eigenvalues to obtain a unique rational expectations equilibrium. In another note we will discuss the issue of determinacy in a simpler version of this model.

DETERMINACY IN SIMPLE NEW KEYNESIAN MODEL

This note discusses the determinacy of the equilibrium of the New Keynesian model with an interest rate rule. Recall the equilibrium conditions of the new Keynesian model that we discussed in a previous note,

$$E_t\left[\hat{y}_{t+1}\right] = \hat{y}_t + \frac{1}{\sigma}\left[\hat{R}_t - E_t\left[\hat{\pi}_{t+1}\right]\right]$$

$$\hat{\pi}_t = \beta E_t\left[\hat{\pi}_{t+1}\right] + \kappa\left[\left(\nu+\sigma\right)\hat{y}_t - \left(1+\nu\right)\hat{A}_t\right]$$

$$\hat{R}_t = \left(1-\rho_R\right)\left[\phi_\pi\hat{\pi}_t + \phi_y\hat{y}_t\right] + \rho_R\hat{R}_{t-1} + v_t$$

$$\hat{A}_{t+1} = \rho_A\hat{A}_t + \varepsilon_{A,t+1}$$

$$v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1}$$

where

$$\kappa = \frac{\theta-1}{\bar{\pi}^2\omega}$$

and ω is the parameter associated with the adjustment cost.

To discuss the issue of determinacy of the equilibrium, we will simplify the model further assuming that $\hat{A}_t = 0$, $v_t = 0$ and $\rho_R = 0$. So basically we consider a model with no smoothing in the interest rate rule. With these assumptions the model can be written as

$$E_t\left[\hat{y}_{t+1}\right] + \frac{1}{\sigma}E_t\left[\hat{\pi}_{t+1}\right] = \hat{y}_t + \frac{1}{\sigma}\hat{R}_t \tag{1}$$

$$\beta E_t\left[\hat{\pi}_{t+1}\right] = \hat{\pi}_t - \gamma\hat{y}_t \tag{2}$$

$$\hat{R}_t = \phi_\pi\hat{\pi}_t + \phi_y\hat{y}_t \tag{3}$$

where we define $\gamma \equiv \kappa\left(\nu+\sigma\right)$.

Replacing the Taylor rule into the first equation gives

$$E_t\left[\hat{y}_{t+1}\right] + \frac{1}{\sigma}E_t\left[\hat{\pi}_{t+1}\right] = \hat{y}_t + \frac{1}{\sigma}\left[\phi_\pi\hat{\pi}_t + \phi_y\hat{y}_t\right]$$

$$E_t\left[\hat{\pi}_{t+1}\right] = \frac{1}{\beta}\hat{\pi}_t - \frac{\gamma}{\beta}\hat{y}_t$$

We can now replace the second equation into the first to get rid of $E_t\left[\hat{\pi}_{t+1}\right]$,

$$E_t\left[\hat{y}_{t+1}\right] + \frac{1}{\sigma}\left[\frac{1}{\beta}\hat{\pi}_t - \frac{\gamma}{\beta}\hat{y}_t\right] = \hat{y}_t + \frac{1}{\sigma}\left[\phi_\pi\hat{\pi}_t + \phi_y\hat{y}_t\right]$$

or

$$E_t\left[\hat{y}_{t+1}\right] = \left(1 + \frac{\phi_y}{\sigma} + \frac{\gamma}{\sigma\beta}\right)\hat{y}_t + \frac{1}{\sigma}\left(\phi_\pi - \frac{1}{\beta}\right)\hat{\pi}_t.$$

We reduced the model to the following system of two equations:

$$\begin{aligned} E_t\left[\hat{y}_{t+1}\right] &= \left(1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right)\right)\hat{y}_t + \frac{1}{\sigma}\left(\phi_\pi - \frac{1}{\beta}\right)\hat{\pi}_t \\ E_t\left[\hat{\pi}_{t+1}\right] &= \frac{1}{\beta}\hat{\pi}_t - \frac{\gamma}{\beta}\hat{y}_t \end{aligned}$$

This can be written in matricial form as

$$E_t\begin{bmatrix} \hat{y}_{t+1} \\ \hat{\pi}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) & \frac{1}{\sigma}\left(\phi_\pi - \frac{1}{\beta}\right) \\ -\frac{\gamma}{\beta} & \frac{1}{\beta} \end{bmatrix}\begin{bmatrix} \hat{y}_t \\ \hat{\pi}_t \end{bmatrix} \tag{4}$$

This system does not have any state variable. Therefore, the Blanchard and Kahn condition implies that, for the equilibrium to be unique, both eigenvalues of the above coefficients matrix must be larger than one in absolute value. If one of the eigenvalues is smaller than one, the model has multiple equilibria.

We now recall two useful facts about eigenvalues of a 2×2 matrix:

1. The product of the eigenvalues is equal to the determinant of the matrix
2. The sum of the eigenvalues is equal to the trace of the matrix.

The first result implies

$$\begin{aligned} \lambda_1\lambda_2 &= \det\begin{bmatrix} 1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) & \frac{1}{\sigma}\left(\phi_\pi - \frac{1}{\beta}\right) \\ -\frac{\gamma}{\beta} & \frac{1}{\beta} \end{bmatrix} \\ &= \left[1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right)\right]\frac{1}{\beta} + \frac{\gamma}{\beta\sigma}\left(\phi_\pi - \frac{1}{\beta}\right) \\ &= \frac{1}{\beta} + \frac{\phi_y}{\sigma\beta} + \frac{\gamma}{\sigma\beta^2} + \frac{\gamma\phi_\pi}{\beta\sigma} - \frac{\gamma}{\beta^2\sigma} \\ &= \frac{1}{\beta} + \frac{\phi_y}{\sigma\beta} + \frac{\gamma\phi_\pi}{\beta\sigma} > 0 \end{aligned} \tag{5}$$

using $\phi_\pi > 0$ and $\phi_y > 0$.

The second result implies

$$\begin{aligned} \lambda_1 + \lambda_2 &= trace \begin{bmatrix} 1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) & \frac{1}{\sigma}\left(\phi_\pi - \frac{1}{\beta}\right) \\ -\frac{\gamma}{\beta} & \frac{1}{\beta} \end{bmatrix} \\ &= 1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) + \frac{1}{\beta} > 0. \end{aligned} \tag{6}$$

From these results it follows that both eigenvalues are positive. We can also prove that they must be real as well. Since both eigenvalues are positive and real, for the equilibrium to be unique it must be the case that both eigenvalues are larger than 1. Therefore, the necessary condition for a unique equilibrium can be written as

$$(\lambda_1 - 1)(\lambda_2 - 1) > 0.$$

Expanding this expression gives

$$\lambda_2\lambda_1 - \lambda_2 - \lambda_1 + 1 > 0$$

or

$$\lambda_1\lambda_2 - (\lambda_1 + \lambda_2) > -1.$$

Using (5) and (6) to replace $\lambda_1\lambda_2$ and $\lambda_1 + \lambda_2$ gives

$$\frac{1}{\beta} + \frac{\phi_y}{\sigma\beta} + \frac{\gamma\phi_\pi}{\beta\sigma} - \left(1 + \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) + \frac{1}{\beta}\right) > -1$$

This implies

$$\frac{1}{\beta} + \frac{\phi_y}{\sigma\beta} + \frac{\gamma\phi_\pi}{\beta\sigma} - 1 - \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) - \frac{1}{\beta} > -1$$

or

$$\frac{\phi_y}{\sigma\beta} + \frac{\gamma\phi_\pi}{\beta\sigma} - \frac{1}{\sigma}\left(\phi_y + \frac{\gamma}{\beta}\right) > 0$$

Multiplying by σ and rearranging the inequality we obtain

$$\phi_y\left(\frac{1}{\beta} - 1\right) + \frac{\gamma\phi_\pi}{\beta} > \frac{\gamma}{\beta}$$

Multiplying by β/γ then gives

$$\phi_y\left(\frac{1-\beta}{\gamma}\right) + \phi_\pi > 1 \tag{7}$$

This is a necessary condition for uniqueness of the equilibrium of the New Keynesian model.

It is a constraints on the parameters of the Taylor rule. The Taylor principle $\phi_\pi > 1$ is a sufficient condition for the equibrium to be determinate. But you can also have $\phi_\pi < 1$ and ϕ_y sufficiently large and the above inequality still hold.