Financial Econometrics - Time Series

Johnew Zhang

March 5, 2018

Contents

1 Basic Time Series Concepts

Suppose stochastic process Y_t. We need to do inference on $\{y_t\}_{t=1}^T$.

Let $f_{Y_t}(y_t)$ unconditional density of Y_t and its mean is

$$E(Y_t) = \int_{-\infty}^{\infty} y_t f_Y(y_t) dy_t$$

where $Y_t = \mu + \varepsilon_t$, $E(\varepsilon_t) = 0, V(\varepsilon_t) = \sigma^2$. It is easy to see $E(Y_t) = \mu$.

Alternatively, there may be period so we can define $Y_t = \beta t + \varepsilon_t$. Here $E(Y_t) = \beta t$ and $V(Y_t) = E((Y_t - \mu_t)^2) = \int_{-\infty}^{\infty}(y_t - \mu_t)^2 f_Y(y_t) dy_t = \gamma_{0t} = \sigma^2$

Definition 1. *Auto-covariance: $\{Y_t\}_{t=1}^T$. Consider vector $x_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-j} \end{pmatrix}$. The joint distribution is $(Y_t, Y_{t-1}, \cdots, Y_{t-j})$. Therefore the jth auto-covariance is*

$$\gamma_{jt} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (y_t - \mu_t)(y_{t-j} - \mu_{t-j}) f_{Y_t, \cdots, Y_{t-j}}(y_t, \cdots, y_{t-j}) dy_t \cdots d_{yt-j}$$

Serial correlation implies non-zero auto-covariances.

Definition 2. *Stationarity:*

- *Covariance Stationarity means $E(Y_t) = \mu$ and $E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j$ (it does not dependent on time t. For scalars, $\gamma_j = \gamma_{-j}$ and for matrix C_j, it will be $C_j = C_j'$.*
- *Strict Stationarity: The entire joint distribution of $(Y_t, \cdots, Y_{t-j_1}, \cdots Y_{t-j_2}, \cdots, Y_{t-j_n})$ depends only on the intervals seperating the dates $(j_1, j_2, \cdots, j_n)$.*
- *Note that a process is strictly stationary with finite second moments, then it must be covariance-stationary. Strict stationarity does not imply weak stationarity (Cauchy distribution). Weak does not imply the strict stationarity. If the processes are Gaussian, then weak is equivalent to the strict.*

Definition 3. *Ergodicity: Time series averages are going to converge to the unconditional moments as $T \to \infty$. It means $\bar{y} = \frac{1}{T}\sum_{t=1}^T y_t \to \mu$ as $T \to \infty$.*

1.1 Some Processes

- ε_t is a white noise if $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = \sigma^2$ and $E[\varepsilon_t \varepsilon_{t-j}] = 0, \forall j \neq 0$.
- Moving average process is defined as $y_t = \varepsilon_t + \theta\varepsilon_{t-1} + \mu$. This is called the first-order MA.

$$E[y_t] = E[\varepsilon_t] + \theta E[\varepsilon_{t-1}] + E[\mu] = \mu$$
$$V(y_t) = E[(y_t - \mu)^2] = E[[\varepsilon_t + \theta\varepsilon_{t-1})^2] = E[\varepsilon_t^2 2\varepsilon_t\varepsilon_{t-1}\theta + \theta^2\varepsilon_{t-1}^2] = (1+\theta^2)\sigma^2$$

Let auto-covariance

$$E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \theta\sigma^2$$

First-order auto-correlation is

$$\frac{\gamma_1}{\sqrt{\gamma_0}\sqrt{\gamma_1}} = \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$

where $\gamma_j = 0, j > 1$

- pth order moving average process

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_p \varepsilon_{t-p}$$

where $E[y_t] = \mu$ and $V(y_t) = E[(y_t - \mu)^2] = E[\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_p \varepsilon_{t-p}]^2 = 1 + \sum_{i=1}^{p} \theta_p^2 \sigma^2$ and $E[\varepsilon_t \varepsilon_{t-j}] = 0$ and $E[\varepsilon_{t-j}^2] = \sigma^2$.

The jth autocovariance is

$$\begin{aligned} \gamma_j &= E[(y_t - \mu)(y_{t-j} - \mu)] = E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_p \varepsilon_{t-p})(\varepsilon_{t-j} + \theta_1 \varepsilon_{t-j-1} + \cdots + \theta_p \varepsilon_{t-j-p}) \\ &= \theta_j \sigma^2 + \theta_{j+1}\theta_1 \sigma^2 + \cdots + \theta_p \theta_{p-j} \sigma^2 \qquad j = 1, \cdots, p \\ &= 0 \qquad j > p \end{aligned}$$

- ∞ order moving average processes

$$y_t = \mu + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}$$

$$V(y_t) = E[(y_t - \mu)^2] = (\sum_{j=0}^{\infty} \psi_j^2) \sigma^2$$

where ψ_j is square summable. If $\sum^{\infty} |\psi_j| < \infty$, then it is ergodic for mean absolute summability.

- Autoregressive Processes: First order AR processes,

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

where $|\phi| < 1$ for covariance stationarity.

Define the lag operator L such that $L y_t = y_{t-1}$. Then $y_t(1 - \phi L) = c + \varepsilon_t$ where $|\phi| < 1$. Therefore

$$y_t = (1 - \phi L)^{-1}(c + \varepsilon_t) = \frac{c}{1 - \phi} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots$$

$$E[y_t] = \frac{c}{1 - \phi}$$

$$V(y_t) = \sigma^2(1 + \phi^2 + \phi^4 + \phi^6 + \cdots) = \frac{\sigma^2}{1 - \phi^2}$$

The first order autocovariance is

$$\begin{aligned} E[(y_t - \mu)(y_{t-1} - \mu)] &= E[\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots)(\varepsilon_{t-1} + \phi \varepsilon_{t-2} + \phi^2 \varepsilon_{t-3} + \cdots) \\ &= \phi \sigma^2 + \phi^3 \sigma^2 + \phi^5 \sigma^2 + \cdots \\ &= \frac{\phi \sigma^2}{1 - \phi^2} \end{aligned}$$

The jth autocovariance is $\frac{\phi^j \sigma^2}{1 - \phi^2}$ and the atuocorrelation is just $\frac{\gamma_j}{\gamma_0} = \phi^j$

For $AR(p) = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$, then

$$y_t(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) = c + \varepsilon_t$$

The p-th order polynomial L is

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z)$$

The roots (λ_j^+) to be outside the unit circle for covariance stationarity.

Variance of $AR(p)$ is

$$E[(y_t - \mu)^2] = \phi_1 [E(y_{t-1} - \mu)(y_t - \mu)] \phi_2 [E(y_{t-2} - \mu)(y_t - \mu)] + \cdots + \phi_p E[(y_{t-p} - \mu)(y_t - \mu)] + E[\varepsilon_t (y_t - \mu)]$$

where $\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2$
Multiply through $y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$ with $y_{t-1} - \mu$ and take $E[]$. Then

$$\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1 + \cdots + \phi_p\gamma_{p-1}$$
$$\vdots$$
$$\gamma_p = \phi_1\gamma_p + \phi_2\gamma_{p-2} + \cdots + \phi_p\gamma_0$$

divide above by γ_0 then we can solve the system of equations so

$$p_j = \phi_1 p_j + \phi_2 p_{j-2} + \cdots + \phi_p p_{j-p}, j > p$$

This processes is called Yule-Walker.

- $ARMA(p,q)$ is defined

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

$ARMA(1,1)$ without a constant term. $y_t(1 - \phi_1 L) = \varepsilon_t(1 + \theta_1 L)$. notice if $\theta_1 = -\phi_1$, lag polynomial cancels.

Definition 4. *Autocovariance generating function is defined as when γ_j is absolutely summable, then*

$$g_y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j$$

for complex scalar z.

The Fourier Transform of a time series $\{x_t\}$ is $x(\omega) = \sum_{t=-\infty}^{\infty} e^{-iwt}x_t$ as a complex function of ω. Here ω is the frequency. The inverse Fourier transformation is

$$x_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}x_t d\omega$$

Hence we can define the Fourier transform of the autocovariance as

$$\begin{aligned} S(\omega) &= \sum_{j=-\infty}^{\infty} e^{-i\omega j}\gamma_j \\ &= \gamma_0(\cos(\omega) + i\sin(\omega)) && \gamma_j = -\gamma_j \\ &+ \sum_{j=1}^{\infty} \cos(\omega j)\gamma_j + i\sin(\omega j)\gamma_j && \cos(x) = \cos(-x), \sin(x) = -\sin(-x) \\ &= \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j\cos(\omega_j) \end{aligned}$$

For auto-correlation,

$$f(\omega) = \frac{S(\omega)}{\gamma_0} = \sum_{j=-\infty}^{\infty} e^{-i\omega j}\rho_j$$

where $\rho_j = \frac{\gamma_i}{\gamma_0}$ The inverse is

$$\rho_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega j} f(\omega)d\omega$$

When $j = 0$,

$$1 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)d\omega$$

Here $\frac{f(\omega)}{2\pi}$ looks like a density function. This is called spectrum density function.

For $MA(1)$,

$$\begin{aligned} g_y(z) &= \theta\sigma^2 z^{-1} + (1+\theta^2)\sigma^2 z^0 + \theta\sigma^2 z^1 \\ &= \sigma^2(\theta z^{-1} + (1+\theta^2) + \theta z) \\ &= \sigma^2(1+\theta z)(1+\theta z^{-1}) \end{aligned}$$

For $MA(q)$,

$$g_y(z) = \sigma^2(1+\theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q)(1+\theta_1 z^{-1} + \theta_2 z^{-2} + \cdots + \theta_q z^{-q})$$

For $AR(1)$,

$$g_y(z) = \frac{\sigma^2}{(1-\phi z)(1-\phi z^{-1})}$$

Definition 5. *Invertibility: ε_t is recoverable from y_t history. Then*

$$y_t = \mu + (1+\theta L)\varepsilon_t$$

If $|\theta| < 1$ we can multiply by $(1+\theta L)^{-1}$.

$$(1 - \theta L + \theta^2 L^2 + \cdots)(y_t - \mu) = \varepsilon_t$$

MA process is invertible.

$$g_y(z) = \sigma^2(1+\theta z)(1+\theta z^{-1})$$

Consider $\tilde{Y}_t$, $(\tilde{Y}_t - \mu) = (1+\tilde{\theta}L)\tilde{\varepsilon}_t$ then

$$\begin{aligned} g_{\tilde{y}}(z) &= \tilde{\sigma}^2(1+\tilde{\theta}z)(1+\tilde{\theta}z^{-1}) \\ &= \tilde{\sigma}^2(\tilde{\theta}z)(\tilde{\theta}^{-1}z^{-1}+1)(\tilde{\theta}z^{-1})(\tilde{\theta}^{-1}z+1) \end{aligned}$$

Let $\theta = \tilde{\theta}^{-1}, \sigma^2 = \tilde{\sigma}^2\tilde{\theta}^2$, $y_t, \tilde{y}_t$ are the same autocovariances and same mean but y_t not invertible so cannot inverse $\tilde{\varepsilon}_t$.

1.2 How should we define market efficiency?

There should have no predictability of returns. Let $p_t = \log S_t$ and $p_{t+1} = p_t + \varepsilon_{t+1} + \mu$. Stock return will be $r_{t+1} = p_{t+1} - p_t = \varepsilon_{t+1} + \mu$ so there is no serial correlation. Early research fail to reject the $\rho = 0$. One way to look at it is to check

$$r_{t+1} = \mu + \rho r_t + \varepsilon_{t+1}$$

run a regression

$$\hat{\rho} = \sum_{t=1}^{T}(r_{t+1} - \bar{r})(r_t - \bar{t}) / \sum_{t=1}^{T}(r_t - \bar{r})^2$$

Under the null hypothesis, $\rho = 0$. Test statistic. $\hat{\rho}/\sqrt{\hat{\sigma}^2/T} \sim N(0,1)$ asymptotically. Under the alternative, $\rho > 0$.

We will do Monte Carlo under the null or alternative with sample size T. Then a histogram can be generated for the test statistic.

(check paper Poterba, Summers) Suppose $p_t = p_t^* + \mu_t$ where u_t is serially correlated and p_t^* is a random walk. Here $r_t = p_t - p_{t-1} = p_t^* - p_{t-1}^* + \mu_t - \mu_{t-1}$ and $u_t = eu_{t-1} + v_t$ where v_t is iid. There will be serial correlation in returns but $r_t = p_t^* - p_{t-1}^* + (\rho - 1)u_{t-1} + v_t$. Suppose $\rho = 0.98$. They set a variance ratio as

$$V(\sum_{k=1}^{24} r_{t+k}) / 2V(\sum_{k=1}^{12} r_{t+k})$$

2 Forecasting

Suppose we want to forecast based on Y_{t+1} and x_t = data. Let $y^*_{t+1,t}$ be forecast. We have quadratic loss function, $E[Y_{t+1} - Y^*_{t+1,t}]^2$. On Hamilton, it proves $E[Y_{t+1}|Y_t]$ is the minimum mean square error. The linear projection is $Y^*_{t+1,t} - x_t'\alpha$. Forecast error is $y_{t+1} - x_t'\alpha$. Linear projection makes forecast error orthogonal to x_t, that is

$$E[x_t(Y_{t+1} - x_t'\alpha)] = 0$$
$$E[x_tY_{t+1}] - E[x_tx_t']\alpha = 0$$
$$\alpha = E[x_tx_t']^{-1}E[x_tY_{t+1}] \qquad \text{population statistics}$$

OLS regression of y_{t+1} on x_t is

$$y_{t+1} = x_t'\beta + u_{t+1}, t = 1, \cdots, T$$

where $b = \left[\sum_{t=1}^T x_tx_t'\right]^{-1}\left[\sum_{t=1}^T x_ty_{t+1}\right]$ as sample estimate.

With covariance stationary, sample moment converges to the population moments as $T \to \infty$.

$$\frac{1}{T}\sum_{t=1}^T x_tx_t' \overset{p}{\to} E[x_tx_t']$$
$$\frac{1}{T}\sum_{t=1}^T x_ty_{t+1} \overset{p}{\to} E[x_tY_{t+1}]$$
$$b \overset{p}{\to} \alpha$$

Here we are assuming data are ergodic for second moments.

2.1 Wold's Decomposition Theorem

Any covariance stationary

$$Y_t = \sum_{j=0}^\infty \psi_j\varepsilon_{t-j} + k_t$$

where k_t is the linearly deterministic. and $\psi_0 = 1, \sum_{j=0}^\infty \psi_j^2 < \infty$ and $\varepsilon_t = y_t - \hat{E}[Y_t|Y_{t-1}, \cdots]$ a linear projection errors.

Suppose we know ε_t's, what is forecast of y_{t+s}?

$$Y_{t+s} = \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + \cdots$$
$$\hat{E}[Y_{t+s}|\varepsilon_t\varepsilon_{t-1}\cdots] = \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \cdots$$

The MSE of forecast is $(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{s-1}^2)\sigma^2$ where $\sigma^2 = V(\varepsilon)$.

Define $\frac{\psi(L)}{L^s} = L^{-s} + \psi_1L^{1-s} + \psi_2L^{2-s} + \cdots$.

Define annihilation operator $[\;]_+$ that sets negative powers to 0.

$$\hat{E}[Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \cdots] = \left[\frac{\psi(L)}{L^s}\right]_+ \varepsilon_t$$

Consider forecast of Y_{t+s} based on $Y_t, Y_{t-1}, \cdots$.

$$\eta(L)Y_t = \varepsilon_t$$
$$\eta(L) = \sum_{j=0}^\infty \eta_jL^j, \eta_0 = 1, \sum_{j=0}^\infty |\eta_j| < \infty$$

for invertible representation $\eta(L) = \psi(L)^{-1}$

$$\hat{E}[Y_{t+s}|Y_t, Y_{t-1}\cdots] = \left[\frac{\psi(L)}{L^s}\right]_+ \eta(L)Y_t = \left[\frac{\psi(L)}{L^s}\right]_+ \frac{1}{\psi(L)}Y_t$$

is called the Weiner-Kolmogorov Prediction Form.

3 Introduction to the Generalized Method Moments (Hayashi)

3.1 Endogeneity Bias

Coffee market with demand $q_t^d = \alpha_0 + \alpha_1 p_t + u_t$ where u_t is the unobservable shifter in demand $\alpha_1 < 0$. Supply is $q_t^s = \beta_0 + \beta_1 p_t + v_t$ where v_t shifts supply. Equilibrium $q_t^d = q_t^s = q_t$. We observe p_t and q_t.

Solution is

$$p_t = \frac{\beta_0 + \alpha_0}{\alpha_1 - \beta_1} + \frac{v_t - u_t}{\alpha_1 - \beta_1}$$

and

$$q_t = \frac{\alpha_1 \beta_0 - d_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha v_t - \beta_1 u_t}{\alpha_1 - \beta_1}$$

p_t increases with $r_t < 0, u_t > 0$ and $\alpha_1 < 0, \beta_1 > 0$.

OLS of q_t on p_t gives you $q_t = \delta_0 + \delta_1 p_t + \varepsilon_t$. Here

$$\hat{\delta}_1 = \frac{cov(q_t, p_t)}{var(p_t)} = \frac{cov(\alpha_0 + \alpha_1 p_t + u_t, p_t)}{var(p_t)}$$
$$= \alpha_1 + \frac{cov(u_t, p_t)}{var(p_t)} \neq 0$$

This is called Endogeneity Bias to OLS or simultaneous Equation Bias. The solution for this dilemma is to estimate the demand curve if we have another variable x_t that shifts the supply curve.

Here we can have $v_t = \beta_2 x_t + \zeta_t$ where ζ_t is a new shock. The new innovations are

$$q_t^d = \alpha_0 + \alpha_1 p_t + u_t$$

$$q_t^s = \beta_0 + \beta_1 p_t + \beta_2 x_t + \zeta_t$$

where $E[x_t \zeta_t] = 0$.

Here

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_t + \frac{\zeta_t - u_t}{\alpha_1 - \beta_1}$$
$$q_t = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} x_t + \frac{\alpha_1 \zeta_t - \beta_1 u_t}{\alpha_1 - \beta_1}$$

where $E[x_t \zeta_t] = 0$ and $E[x_t u_t] = 0$. The above solution is called the reduced form simultaneous equation system.

Express endogenous variables in terms of exogenous variables. OLS of p_t on x_t with $\hat{\delta}_1 = \frac{\beta_2}{\alpha_1 - \beta_1}$. OLS of q_t on x_t with $\hat{\delta}_2 = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1}$. Thus

$$\frac{\hat{\delta}_2}{\hat{\delta}_1} = \alpha_1$$

This estimator is called the instrumental variables estimator with x_t as the instrument.

3.2 Two-Staged Least Square

1. Regress p_t on x_t with OLS. Then

   $$\hat{p}_t = \hat{\pi}_0 + \hat{\pi}_1 x_t$$

   The fitted value is

   $$p_t = \hat{p}_t + \text{error}_{\text{ at lagged}+\sigma \hat{p}_t}$$

2. Regress q_t on $\hat{p}_t$.

   $$q_t = \alpha_0 + \alpha_1 \hat{p}_t + u_t + \alpha_1 (p_t - \hat{p}_t)$$

   where the last two terms is the composite error and is orthogonal to $\hat{p}_t$ so the OLS of q_t on $\hat{p}_t$ gives

   $$\hat{\alpha}_1 = \frac{cov(q_t, \hat{p}_t)}{var(p_t)} = \alpha_1$$

The fundamental equation of finance tells

$$E_t[m_{t+1}R_{t+1}] = 1$$

Can we estimate the parameters in this equation?

3.3 Single Equation GMM

Suppose

$$y_t = z_t'\delta + \varepsilon_t, t = 1, \cdots, T$$

where z_t is $L \times 1$.

A3.1 : Linearity

A3.2 : Ergodic stationarity such that x_t is a $k \times 1$ vector of instruments and w_t is unique elements of $(y_t, z_t', x_t')'$. This is stationary and ergodic.

A3.3 $E[x_t\varepsilon_t] = 0$ is the orthogonality conditions. Let's define $g_t = x_t\varepsilon_t = x_t(y_t - z_t'\delta)$ is the function of data and parameters. Here the variables are $k \times 1$.

A4.4 $k \geq L$ is the rank condition for identification. Here $E[x_t z_t']$ is full column rank. Weak instruments satisfy this axiom poorly. If identification, $E[g_t(w_t, \delta_0)] = 0$ at the true δ_0. It is not 0 at $\delta \neq \delta_0$.

$$E[x_t(y_t - z_t'\delta)] = 0$$

$$\sigma_{xy} - \Sigma_t'\delta = 0$$

in terms of population parameters. This is a system of k equations in $L \leq k$ unknowns. The necessary and sufficient condition for one solution is $k \geq L$. Over-identification is $k > L$. Just or exact is $k = L$ and under-identification is $k < L$.

A3.5 g_t is a Martingale difference sequence with finite second moments. g_t is a Martingale difference sequence if

$$E[g_t | g_{t-1} g_{t-2} \cdots] = 0$$

x_t is a Martingale such that

$$E[x_t | \Phi_{t-1}] = x_{t-1}$$

If Φ_{t-1} is x_{t-1}, x_{t-2},

$$s_t = \sum_{j=1}^{t} g_{t-j} = g_t + g_{t-1} + g_{t-2} + \cdots = g_t + s_{t-1}, E[s_t] = E[g_t] + s_{t-1} = s_{t-1}$$

s_t is a Martingale and $g_t = s_t - s_{t-1}$ is Martingale difference sequence.

A3.6 Finite 4th moment of the w_t process

If g_t is a Martingale difference sequence (MDS), then $E[g_t g_t']$ = variance of $g_t = S$. Billingsley (1961) used Central Limit Theorem for MDS. if g_t is MDS that is stationary and ergodic with $E[g_t g_t'] = S$ then $\bar{g} = \frac{1}{T}\sum_{t=1}^{T} g_t$ is sample mean and $\sqrt{T}\bar{g} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t \xrightarrow{d} N(0, S)$.

Comments

- If instruments include a constant $E[\varepsilon_t] = 0$.
- Alternative to A3.5 is $E[\varepsilon_t | x_t, x_{t-1}, \cdots] = 0$.
- $g_t g_t' = \varepsilon_t^2 x_t x_t'$.
- We will relax the linearity and serial correlation of g_t.

GMM defined an Economic model that gives a set of theoretical orthogonality condition $E[x_t\varepsilon_t] = 0$. This is the population moments. GMM chooses parameters to set a weighted average of sample moments as close to zero as possible. (corresponding to the population moments). The model says

$$g_T(\tilde{\delta}) = \frac{1}{T}\sum_{t=1}^{T} g_t(w_t, \tilde{\delta}_t)$$

where $E[g_t(\delta_0)] = 0$. Here

$$g_T(\tilde{\delta}) = \frac{1}{T}\sum_{t=1}^{T} x_t y_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_t z_t'\right)\tilde{\delta} = s_{xy} - S_{xz}\tilde{\delta}$$

If $k = L$, just-identified, what is $\hat{\delta}$? $\hat{\delta}_{xz} = S_{xz}^{-1}s_{xy}$. Sets the sample orthogonality conditions to zero. If $x_t = z_t$, then this is OLS.

If $k > L$, over-identified, GMM objective function

$$J(\tilde{\delta}, W) = T g_T(\tilde{\delta})' W g_T(\tilde{\delta})$$

where g_T is the sample mean. Choose $\hat{\delta}$ as argmin of $J(\tilde{\delta}, W)$. That is

$$J(\tilde{\delta}, W) = T(s_{xy} - S_{xz}\tilde{\delta})' W (s_{xy} - S_{xz}\tilde{\delta})$$

minimized by choice of $\tilde{S}$. FOC:

$$S_{xz}' W s_{xy} - S_{xz}' W S_{xz}\hat{\delta} = 0$$
$$\hat{\delta} = \left[S_{xz}' W S_{xz}\right]^{-1} S_{xz}' W s_{xy}$$

Single equations GMM estimator with instrumental variable. Here W must be positive-definite.

$$s_{xy} = \frac{1}{T}\sum_{t=1}^{T} x_t y_t = \frac{1}{T}\sum_{t=1} T x_t(z_t'\delta_0 + \varepsilon_t) = S_{xz}\delta_0 + \frac{1}{T}\sum_{t=1}^{T} x_t\varepsilon_t = S_{xz}\delta_0 + g_T(\delta_0)$$
$$\hat{\delta} = \left(S_{xz}' W S_{xz}\right)^{-1} S_{xz}' W (S_{xz}\delta_0 + g_T)$$
$$\hat{\delta} = \delta_0 + \left(S_{xz}' W S_{xz}\right)^{-1} S_{xz}' W g_T$$
$$\sqrt{T}(\hat{\delta} - \delta_0) = \left(S_{xz} W S_{xz}\right)^{-1} S_{xz}' W \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t$$

converges to $N(0, S)$ and $S = E[g_t g_t']$, the variance of g_t. As $T \to \infty$, sample moments $S_{xz} \xrightarrow{p} \Sigma_{xz}$.

$$Avar(\hat{\delta}) = \left(\Sigma_{xz}' W \Sigma_{xz}\right)^{-1} \Sigma_{xz}' W S W \Sigma_{xz} \left(\Sigma_{xz}' W \Sigma_{xz}\right)^{-1}$$

Estimator of Avar use S_{xz} for Σ_{xz}., we need. to estimate $\hat{S}$ for S. That is

$$\hat{S} = \frac{1}{T}\sum_{t=1}^{T} \hat{g}_t \hat{g}_t' = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon_t}^2 x_t x_t'$$

$$\hat{\varepsilon_t} = y_t - z_t'\hat{\delta} = z_t' S_t + \varepsilon_t - z_t'\hat{\delta}$$
$$= \varepsilon_t + z_t'(\delta_0 - \delta)$$

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 \sum_{t=1}^{T}\left(\varepsilon_t^2 - 2(\hat{\delta} - \delta_0)' z_t \varepsilon_t + (\hat{\delta} - \delta_0\right)' z_t z_t' (\hat{\delta} - \delta_0)$$

as $T \to \infty$, then $\frac{1}{T}\varepsilon_t^2 \to E[\varepsilon_t^2], \hat{\delta} - \delta_0 \to 0$. $\frac{1}{T}\sum_{t=1}^{T} z_t\varepsilon_t \to$ finite middle $\to 0$ and $\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \to$ finite.

To test this, we check $\sqrt{T}(\hat{\delta}-\delta_0) \xrightarrow{d} N(0, A\hat{var}(\hat{\delta}))$. Test $\hat{\delta}_2 = \delta_{1,0}$. First we will check

$$\frac{\hat{\delta}_2 - \delta_{1,0}}{se(\hat{\delta}_2)}$$

where

$$se(\hat{\delta}_2) = \sqrt{e_i' Avar(\hat{\delta}) e_i / T}$$

and $e_i = [0, 0, \cdots, 1, \cdots]$, the ith element is 1.

Robust to conditional Heteroskedasticity. Wald Test of a vector of linear restriction is

$$H_0 : R\delta_0 = r$$

where r is number of restriction. Then

$$\sqrt{T}(R\hat{\delta} - R\delta_0) \xrightarrow{d} N(0, RA\hat{var}(\hat{\delta})R')$$

where

$$Wald = T(R\hat{\delta} - r)' \left[RA\hat{var}(\hat{\delta})R'\right]^{-1} g(R\hat{\delta} - r)$$

Non-linear restriction

$$H_0 : a(\delta_0) = 0$$
$$A(\delta) = \Delta_\delta a(\delta)$$

The Wald test is

$$Ta(\delta)' \left\{A(\hat{\delta})A\hat{var}(\hat{\delta})A(\hat{\delta})'\right\}^{-1} a(\hat{\delta})$$

where

$$a(\hat{\delta}) = a(\delta_0) + A(\bar{\delta})(\hat{\delta} - \delta_0)$$
$$\sqrt{T}a(\hat{\delta}) \to A(\bar{\delta})\sqrt{T}(\hat{\delta} - \delta_0)$$

What is W? Efficient GMM uses $W = S^{-1}$.

$$Avar(\hat{\delta}) = \Sigma_{xz}^{-1} S^{-1} \Sigma_{xz}^{-1}$$

.

Hanson 1982 his theorem (3.2) proves that this is the smallest asymptotic variance of $\hat{\delta}$ for orthogonality conditions. (Hysashi P245 Prob 3).

However, we don't know S. There is a 2-step efficient GMMs:

1. Use known W where $W = I_k$. Hisashi recommends to use $W = S_{xx}^{-1}$. If this is used, then

$$\hat{\delta}_1 = (S_{xz}' S_{xx}^{-1} S_{xz})^{-1} S_{xz}' S_{xx}^{-1} s_{xy}$$

$$\hat{\varepsilon}_t = y_t - z_t' \hat{\delta}_1$$

2. Use $\hat{\varepsilon}_t$ to estimate

$$\hat{S}_1 = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 x_t x_t'$$

3. Use $W = \hat{S}_1^{-1}$ to estimate

$$\hat{\delta}_2 = (S_{xz}' \hat{S}_1^{-1} S_{xz})^{-1} S_{xz}' \hat{S}_1^{-1} s_{xy}$$

4. Either stop or use $(S_{xz}' \hat{S}_1^{-1} S_{xz})^{-1}$ as $A\hat{var}$ of $\hat{\delta}_2$ and then iterate util convergence. MC process suggests this is suggested but not required.

   Limit the number of orthogonality conditions: distinct elements S are $\frac{k(k+1)}{2}$ additional parameters. T observations and k series can implode the equations very quickly.

3.4 Hansen's J-Test

Model may give overidentify restrictions because the number of orthogonality conditions is greater than the number of parameters. The GMM objective function

$$J(\hat{\delta}, \hat{S}^{-1}) = Tg_T(\tilde{\delta})'\hat{S}^{-1}g_T(\tilde{\delta})$$

We will take the arg min of δ as the above. Suppose this is δ_0. Then

$$J = \sqrt{T}g_T(\delta_0)'S^{-1}\sqrt{T}g_T(\delta_0) \to \chi^2(k)$$

because $\sqrt{T}g_T(\delta_0) \to N(0, S)$. The estimation of $\hat{\delta}$ sets the L linear combination of $\sqrt{T}g_T(\delta) = 0$. We know that $J(\hat{\delta}, \hat{S}^{-1}) \to \chi^2(k - L)$ and this is called the Hansen's J-Test.

$$\begin{aligned}
\sqrt{T}g_T(\hat{\delta}) &= \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} x_t(y_t - z_t'\hat{\delta}) \\
&= \sqrt{T}(s_{xy} - S_{xz}\hat{\delta}) = \sqrt{T}\left[s_{xy} - S_{xz}(S_{xz}'\hat{S}^{-1}S_{xz})^{-1}S_{xz}'\hat{S}^{-1}s_{xy}\right] \\
&= \sqrt{T}\left[I - S_{xz}(S_{xz}'\hat{S}^{-1}S_{xz})^{-1}S_{xz}'\hat{S}^{-1}\right]s_{xy} \\
&= \sqrt{T}\hat{B}s_{xy} && \hat{B} \text{ is not full column rank} \\
& && \hat{B}S_{xz} = 0
\end{aligned}$$

3.5 Likelihood-Ratio Test of H_0

Here H_0 has restrictions on the parameters.

1. Estimate without restrictions and get $\hat{S}_1$. $Tg_T(\hat{\delta})'\hat{S}_1g_T(\hat{\delta})$ where $\hat{\delta}$ is the unrestricted estimators.
2. Estimate with H_0 restrictions using $\hat{S}_1$. $Tg_T(\bar{\delta})'\hat{S}_1g_T(\bar{\delta})$ where $\bar{\delta}$ is the restricted estimators.

Since we know the optimization with the constraints will be larger than the one without so we will check the difference

$$J(\bar{\delta}, \hat{S}_1) - J(\hat{\delta}, \hat{S}_1) \to \chi^2(r)$$

3.6 Newey-West Motivation

Suppose we have $\{y_t\}_{t=1}^T$ n dimensional and covariance stationary. The mean is $E[y_t] = \mu$. Therefore $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T y_t$. $E[\hat{\mu}] = \frac{1}{T}\sum_{t=1}^T E[y_t] = \frac{T}{T}\mu = \mu$ is unbiased.

The variance is

$$\begin{aligned}
E[(\hat{\mu} - \mu)(\hat{\mu} - \mu)'] &= E\left[\frac{1}{T}\sum_{t=1}^{T}(y_t - \mu)\frac{1}{T}\sum_{t=1}^{T}(y_t - \mu)'\right] \\
&= \frac{1}{T^2}E\left[(y_1 - \mu)\sum_{t=1}^{T}(y_t - \mu)' + (y_2 - \mu)\sum_{t=1}^{T}(y_t - \mu)' + \cdots + (y_T - \mu)\sum_{t=1}^{T}(y_t - \mu)'\right] \\
&= \frac{1}{T^2}\left\{T\Gamma_0 + (T-1)[\Gamma_1 + \Gamma_1'] + \cdots + [\Gamma_{T-1} + \Gamma_{T-1}']\right\} \\
TV(\hat{\mu}) &= \Gamma_0 + \frac{T-1}{T}[\Gamma_1 + \Gamma_1'] + \cdots + \frac{1}{T}[\Gamma_{T-1} + \Gamma_{T-1}'] \\
\lim_{T\to\infty} TV(\hat{\mu}) &= \sum_{j=-\infty}^{\infty}\Gamma_j
\end{aligned}$$

Then we know

$$TE\left[(\hat{\mu} - \mu)(\hat{\mu} - \mu)'\right] = \lim_{T\to\infty} E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(y_t, \mu)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(y_t, \mu)\right)'\right]$$

Then the variance of $\sqrt{T}g_T$ is $S = \sum_{j=-\infty}^{\infty} \Gamma_j$.

Recall $S(\omega) = \sum_{j=-\infty}^{\infty} \frac{1}{2\pi}\Gamma_j e^{-i\omega j}$ with the spectrum density at frequency ω. The above condition is just $S(0)$.

3.7 Hansen-Hodrick JPE(1980)

Forward exchange rates as predictors of Future spot rates.

$$F_{t,k} = E_t[S_{t+k}]$$

where

$$S_{t+k} = E_t[S_{t+k}] + \varepsilon_{t,t+k}, \text{ with some reaction to news}$$

However the actual data is not very stationary so the paper propose using the rates of appreciation $s_{t+k} - s_t$ (i.e. .05 means 5% appreciations in dollar) in logs and forward premium $f_{t,k} - s_t$ in logs (i.e. .02 means 2% more expensive to purchase ponders with dollars for delivery in k periods).

With rational expectation that

$$s_{t+k} - s_t = E_t(s_{t+k} - s_t) + u_{t+k,t}$$

and $E_t(u_{t+k,t}) = 0$, under null hypothesis

$$E_t[s_{t+k} - s_t] = \alpha + (f_{t,k} - s_t)$$

Alternatively

$$s_{t+k} - s_t = \alpha + \beta(f_{t,k} - s_t) + u_{t+k,t}$$

where $\beta = 1$ as null is in interest. What are legitimate instruments to use? Anything is in the information can be used as the instrument. (e.g., constant, forward premium). The orthogonality condition is

$$E\left[u_{t+k,t}\begin{pmatrix}1\\ f_{t,k} - s_t\end{pmatrix}\right] = 0$$

Then

$$g_t(\delta) = \{(s_{t+k} - s_t) - \alpha - \beta(f_{t,k} - s_t)\}\begin{pmatrix}1\\ f_{t,k} - s_t\end{pmatrix}$$

where $\delta = (\alpha, \beta)'$. Let $y_{t+k} = s_{t+k} - s_t$, $x_t = \begin{pmatrix}1\\ f_{t,k} - s_t\end{pmatrix}$, $y = (y_{1+k}, \cdots, y_{t+k})'$, $X = \begin{pmatrix}x_1'\\ \vdots\\ x_T'\end{pmatrix}$ and $u = \begin{pmatrix}u_{1+k,1}\\ \vdots\\ u_{T+k,T}\end{pmatrix}$. We have $g_T(\delta) = \frac{1}{T}X'\mu$. Based on GMM, we have

$$J(\hat{\delta}, W) = Tg_T(\delta)'Wg_T(\delta) = T\left[\frac{1}{T}X'(y - X\delta)\right]' W\left[\frac{1}{T}X'(y - X\delta)\right]$$

$\hat{\delta} = (X'X)^{-1}X'y$ is OLS.

$$\hat{\delta} = (X'X)^{-1}X'(X\delta_0 + u) = \delta_0 + \left(\frac{(X'X)}{T}\right)^{-1} g_T(\delta_0)$$

$$\sqrt{T}(\hat{\delta} - \delta_0) = \left(\frac{(X'X)}{T}\right)^{-1}\sqrt{T}g_T(\delta_0)$$

$$\sqrt{T}g_T(\delta_0) \to N(0, S)$$

$$S = \sum_{j=-\infty}^{\infty} \Gamma_j, \text{ if } \Gamma_j \neq 0$$

$$\Gamma_j = E[u_{t+k,t} x_t u_{t+k-j,t-j} x'_{t-j}]$$

for $j < k$, $\Gamma_j \neq 0$, $j \geq k$, $\Gamma_j = 0$. The paper was able to sample data more timely than the forecasting. Hansen-Hodrick GMM uses $\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{k-1}(\hat{\Gamma}_j + \hat{\Gamma}'_j)$. Sometimes this estimator does not turns out to be positive definite so Newey-West comes along.

3.8 Non-linear GMM: Consumption-based Asset Pricing

Let $p_{jt} =$ real price of asset j, $d_{jt} =$ real dividend of asset j. $u'(c_t) =$ marginal utility of consumption. The first order condition for equilibrium investment in an asset is the marginal cost is equal to the expected marginal utility in the future

$$u'(c_t)p_{jt} = E_t[\beta u'(c_{t+1})(p_{j,t+1} + d_{j,t+1})], j = 1, \cdots, N \tag{1}$$

One utility function people use is $u(c_t) = \frac{c_t^{1-\alpha}}{1-\alpha}$ (CRRA). Then

$$r_{j,t+1} = \text{ real return } = \frac{p_{j,t+1} - d_{j,t+1}}{p_{jt}}$$

We can divide equation (1) by $u'(c_t)p_{it}$ and take unconditional expectation

$$1 = E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha} R_{j,t+1}\right]$$

must hold for $j = 1, \cdots, N$

Orthogonality condition is

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha} R_{j,t+1} - 1\right] = 0$$

where $\theta = (\alpha, \beta)'$.

$$\varepsilon_{t+1}\left(\theta, R_{t+1}, \frac{c_{t+1}}{c_t}\right) = \left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha} R_{j,t+1} - 1\right]$$

$$E_t\left[\varepsilon_{t+1}\left(\theta, R_{t+1}, \frac{c_{t+1}}{c_t}\right)\right] = 0$$

$$E_t\left[\varepsilon_{t+1}\left(\theta, R_{t+1}, \frac{c_{t+1}}{c_t}\right) \otimes x_t\right] = 0, x_t \in \Phi_t$$

(M instruments usual 1 of which is a constant.)

Define

$$g_t(\theta, w_{t+1}) = \varepsilon_{t+1}\left(\theta, R_{t+1}, \frac{c_{t+1}}{c_t}\right) \otimes x_t$$

where w_{t+1} unique elements of data

Here $E[g_t(\theta, w_{t+1})] = 0$. $g_t(\theta, w_{t+1})$ is a $k = MN$ dimensional time series function of data and parameters.

A1 w_{t+1} is stationary and ergodic. Then, when $g_t(\theta, w_{t+1})$ is continuous in θ for all w_{t+1} and differentiable with respect to θ then

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta, w_{t+1}) \xrightarrow{p} E(g_t(\theta, w_{t+1}))$$

this is the sample mean of the orthogonality condition.

$$G_T(\theta) = \nabla g_T(\theta) \to E[G(\theta)]$$

A2 Identification case: $E(g_t(\theta, w_{t+1})) \neq 0, \forall \theta \neq \theta_0$. Otherwise, 0.

A3 $\sqrt{T} g_T(\theta_0) \xrightarrow{d} N(0, S)$. $S = \sum_{j=-\infty}^{\infty} \Gamma_j$ but theory will often limit j.

GMM objective function is

$$J_T(\hat{\theta}) = \arg\min_{\hat{\theta}} T g_T(\theta)' W g_T(\theta)$$

for some positive definite symmetric $k \times k$ weighting matrix W. For over-identified $k > p$, the FOC is

$$G_T(\hat{\theta}) W g_T(\hat{\theta}) = 0$$

p linear combinations of sample average orthogonality conditions are zero.

$$a_T g_T(\theta) = 0$$

(Hensen and Cochrarne use this) where $a_T = G_T(\hat{\theta})' W$

Apply the mean-value theorem,

$$g_T(\hat{\theta}) = g_T(\theta_0) + G_T(\bar{\theta})(\hat{\theta} - \theta_0)$$

We will substitute into the FOC.

$$G_T(\hat{\theta}) W [g_T(\theta_0) + G_T(\bar{\theta})(\hat{\theta} - \theta_0)] = 0$$

$$\sqrt{T}(\hat{\theta} - \theta_0) = -[G_T(\hat{\theta})' W G_T(\bar{\theta})]^{-1} G_T(\hat{\theta})' W \sqrt{T} g_T(\theta_0)$$

Under the standard regularity conditions,

$$G_T(\hat{\theta}), G_T(\bar{\theta})$$

converges to $E[G(\theta_0)]$.

$$\sqrt{T} g_T(\theta_0) \xrightarrow{d} N(0, S)$$

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, Avar(\hat{\theta}))$$

where $Avar(\hat{\theta}) = (G'WG)^{-1} G'WSWG(G'WG)^{-1}$ and S is asymptotic variance of $g_T(\theta, w_{t+1})$

Setting $W = S^{-1}$ is optimal

$$Avar(\hat{\theta}) = (G'S^{-1}G)^{-1}$$

1. Calculate $\hat{\theta}_1$ with known $W = I$
2. Calculate $\hat{S}_1$ using $\hat{\theta}_1$ to get the variance of $g_t(\hat{\theta}, w_{t+1})$. Impose the lag restrictions on $\hat{\Gamma}_j = 0$.
3. Use $W = \hat{S}^{-1}$ to get $\hat{\theta}_2$ either stop or iterate to convergence.
4. Form $G_T(\hat{\theta}) = \nabla_\theta g_T(\hat{\theta})$ either analytically or numerically. Define a procedure that calculates $g_T(\hat{\theta})$. Taking numerical gradient at $\hat{\theta}$ of procedure.
5. Do tests with
$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \left[G_T(\hat{\theta})' \hat{S}^{-1} G_T(\hat{\theta})\right]^{-1})$$

Suppose we have n assets.

$$\varepsilon_{t+1} = \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\alpha} R_{t+1} - 1$$

$$g_t(\theta, w_{t+1}) = \varepsilon_{t+1}$$

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} g_t(\theta, W_{t+1})$$

where $\theta = (\beta, \alpha)'$

$$G_T(\hat{\theta}) = \nabla_\theta g_T(\hat{\theta}) = \left[\frac{1}{T} \sum_{t=1}^{T} \left(\frac{c_{t+1}}{c_t} \right)^{-\alpha} R_{t+1}; \frac{1}{T} \sum_{t=1}^{T} -\beta \log \left(\frac{c_{t+1}}{c_t} \right) \left(\frac{c_{t+1}}{c_t} \right)^{-\alpha} R_{t+1} \right]$$

$E_t[\varepsilon_{t+1}] = 0$, by theory

$$\hat{S} = \frac{1}{T} \sum_{t=1}^{T} g_t(\hat{\theta}, w_{t+1}) g_t(\hat{\theta}, w_{t+1})'$$

$$\sqrt{T} \left[\begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix} - \begin{pmatrix} \hat{\beta}_0 \\ \hat{\alpha}_0 \end{pmatrix} \right] \xrightarrow{d} N(0, (G_T(\hat{\theta})' \hat{S}^{-1} G_T(\hat{\theta}))^{-1})$$

3.9 The Asymptotic Distribution of the Orthogonality Conditions

$$g_T(\hat{\theta}) = g_T(\theta_0) + G_T(\bar{\theta})(\hat{\theta} - \theta_0)$$

The following is true because the first-order condition

$$G_T' W g_T(\hat{\theta}) = 0 = G_T' W g_T(\theta_0) + G_T' W G_T(\hat{\theta} - \theta_0)$$

We argue that

$$\hat{\theta} - \theta_0 = -(G_T' W G_T)^{-1} G_T' W g_T(\theta_0)$$

$$\begin{aligned} g_T(\hat{\theta}) &= g_T(\theta_0) - G_T(G_T' W G_T)^{-1} G_T' W g_T(\theta_0) \\ &= [I - G_T(G_T' W G_T)^{-1} G_T' W] g_T(\theta_0) \end{aligned}$$

We know that

$$\sqrt{T}(g_T(\theta_0)) \xrightarrow{d} N(0, S)$$

$$\sqrt{T} g_T(\hat{\theta}) \xrightarrow{d} N(0, [I - G(G'WG)^{-1}G'W] S [I - G(G'WG)^{-1}G'W]')$$

If $W = S^{-1}$, then above is going to be reduced to

$$\begin{aligned} &[IG(G'S^{-1}G)^{-1}G'S^{-1}] S [IG(G'S^{-1}G)^{-1}G'S^{-1}]' \\ &= S - G(G'S^{-1}G)^{-1}G'S^{-1}S - G(G'S^{-1}G)^{-1}G'S^{-1} + G(G'S^{-1}G)^{-1}G'S^{-1} \\ &= [S - G(G'S^{-1}G)^{-1}G' \end{aligned}$$

Asymptotically,

$$\sqrt{T} g_T(\hat{\theta}) \xrightarrow{d} N(0, \hat{S} - G_T(G_T' \hat{S}^{-1} G)^{-1} G_T')$$

From Lemma 4.2 (either Hayashi or Hamilton), we know that

$$T g_T'(\hat{\theta}) S^{-1} g_T(\hat{\theta}) \sim \chi^2(r - p)$$

where r is the number equations and p is number of parameters.

$\hat{S}$ estimates the variance of $g_t(\theta)$ where $V(g_t(\theta)) = E[g_t(\theta) g_t(\theta)']$. Under null we have $E[g_t(\theta)] = 0$. Here we have a few tips as given in the following to improve our test

1. Consider the estimate for S as $\frac{1}{T} \sum_{t=1}^{T} g_t(\hat{\theta}) g_t(\hat{\theta})'$ where $g_t(\theta)$ is serially uncorrelated. We can improve the power of the test by setting

$$\hat{S} = \frac{1}{T} \sum_{t=1}^{T} [g_t(\hat{\theta}) - g_T(\hat{\theta})][g_t(\hat{\theta}) - g_T(\hat{\theta})]'$$

2. Scale data so variances of $g_t(\theta)$ are similar.

3. Keep model relatively small. Since $\frac{K(K+1)}{2}$ in S to be unknown, the size of S can be very large.

4. In our asset pricing model

$$E_t[m_{t+1}(\theta) R_{t+1}] = 1$$

   If we have instrument 1 and x_t, then the orthogonality conditions of our model becomes

$$E \left[m_{t+1}(\theta) R_{t+1} \otimes \begin{pmatrix} 1 \\ x_t \end{pmatrix} - 1 \otimes \begin{pmatrix} 1 \\ x_t \end{pmatrix} \right] = 0$$

3.10 Hansen-Hodrick (1983)

Consider conditional CAPM with constant βs with excess return

$$E_t(R_{it+1}) = \beta_i E_t(R_{mt+1}), i = 1, \cdots, N$$

We know from rational expectation

$$R_{it+1} = E_t(R_{it+1}) + \varepsilon_{it+1}, \varepsilon_{it+1} \perp \Phi_t$$

where Φ_t is the information set of the investors. Consider the linear projection of $E_t[R_{mt+1}]$ on to x_t observable. Then

$$E_t(R_{mt+1}) = \alpha + \delta' x_t + v_t, v_t \perp (1, x_t)$$

where this is a latent variable approach.

$$R_{it+1} = \beta_i(\alpha + \delta' x_t) + \beta_i v_t + \varepsilon_{it+1}, i = 1, \cdots, N$$

Let's call

$$u_{it+1} = \beta_i v_t + \varepsilon_{it+1} \perp \begin{pmatrix} 1 \\ x_t \end{pmatrix}$$

Normalize $\beta_1 = 1$. Then we can write the following

$$\begin{pmatrix} R_{it+1} \\ \vdots \\ R_{Nt+1} \end{pmatrix} = \begin{pmatrix} 1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} (\alpha + \delta' x_t) + \begin{pmatrix} u_{1t+1} \\ u_{2t+1} \\ \vdots \\ u_{Nt+1} \end{pmatrix}$$

where our orthogonality conditions are

$$E\left[\begin{pmatrix} u_{1t+1} \\ u_{2t+1} \\ \vdots \\ u_{Nt+1} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ x_t \end{pmatrix}\right] = 0$$

Here we assume x_t has m elements and k from α and δ, $(N01)$ from β's. Then We have $Nk > N - 1 + k$ over-identified GMM.

4 Vector Auto-regression

4.1 Maximum Likelihood Estimation

Let $f(y_t|x_t, y_{t-1}; \theta)$ probability density function of y_t given past x_t and y_{t-1} (the past history). View $f(y_t|x_t, y_{t-1}; \theta)$ as a function of unknown θ and a likelihood function. Here we know

$$\int_A f(y_t|x_t, y_{t-1}; \theta) dy_t = 1$$

Cremer (1946) says under appropriate regularity conditions, we can differentiate the above with respect to θ

$$\int_A \frac{\partial f(y_t|x_t, y_{t-1}; \theta)}{\partial \theta} = 0$$

Multiply by $\frac{f}{f}$ on both side. Then we have

$$\int_A \frac{\partial f(y_t|x_t, y_{t-1}; \theta)}{\partial \theta f(y_t|x_t, y_{t-1}; \theta)} f(y_t|x_t, y_{t-1}; \theta) = 0$$

Thus

$$E\left[\frac{\partial \log f(y_t|x_t, y_{t-1};\theta)}{\partial \theta}\right] = 0$$

Define $s_t(\theta) = \frac{\partial \log f(y_t|x_t,y_{t-1};\theta)}{\partial \theta}$ is the t-th score function. The maximum likelihood function tells us

$$E_{t-1}[s_t(\theta)] = 0$$

and

$$E[s_t(\theta)] = 0$$

Hence the maximum likelihood function is GMM on the score function.

$$L(y_t) = \prod_{t=1}^{T} f(y_t|x_t, y_{t-1};\theta)$$

since the innovations (the residuals basically) in y_t are serially uncorrelated. Then

$$l(y_t) = \sum_{t=1}^{T} \log f(y_t|x_t, y_{t-1};\theta)$$

From maximum likelihood, we have

$$\max l(y_t)$$

to set

$$\frac{1}{T}\sum_{t=1}^{T} s_t(\theta) = \frac{1}{T}\sum_{t=1}^{T} s_t(\theta_0) + \frac{1}{T}\sum_{t=1}^{T}(s_t(\theta) - s_t(\theta_0) = 0$$

We know

$$\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T} s_t(\theta_0)) \xrightarrow{d} N(0, S)$$

where $S = E[s_t(\theta_0)s_t(\theta_0)']$ because $s_t(\theta_0)$ is serially uncorrelated. We know

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial s_t(\theta)}{\partial \theta} \xrightarrow{d} E\left[\frac{\partial^2 \log f}{\partial \theta \partial \theta'}\right] = -G$$

$$\sqrt{T}(\theta - \theta_0) \xrightarrow{d} N(0, G^{-1}SG^{-1})$$

where G has same square dimension as S. In MLE, $S = G = I =$ fisher's information matrix

$$\sqrt{T}(\theta - \theta_0) \xrightarrow{d} N(0, I^{-1})$$

$$S = E\left(\frac{\partial f}{\partial \theta}\frac{\partial f'}{\partial \theta}\right) = -E\left[\frac{\partial^2 f}{\partial \theta \partial \theta'}\right]$$

if the model is true.

4.2 Vector Auto-regression

We will be doing first-order vector-autoregression

$$y_t = Ay_{t-1} + \varepsilon_t$$

with zero means and

$$\underset{N\times 1}{y_t} = \underset{N\times 1}{C} + \underset{N\times N}{\Phi_1} y_{t-1} + \cdots + \Phi_p y_{t-p} + \underset{N\times 1}{\varepsilon_t}$$

where ε_t is serially uncorrelated and $N(0, \Omega)$ and y_t has dimension N. The key is VAR completely characterizes the auto-correlated y_t.

$T+p$ observations (conditional on the first p observation) on y_t. Goal is estimate

$$\theta = (C, \Phi_1, \cdots, \Phi_p, \Omega)$$

Conditional distribution of y_t given in past data is

$$y_t \sim N(C + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p}, \Omega) = N(\Pi' x_t, \Omega)$$

Define

$$x_t = \begin{pmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}, \Pi' = (C, \Phi_1, \cdots, \Phi_p)$$

$$y_{jt} = C_j + \Phi'_{1j} y_{t-1} + \cdots + \Phi'_{pj} y_{t-p} + \varepsilon_{jt}$$

The conditional density function of the t-th observation is going to be

$$f(y_t|x_t, \theta) = (2\pi)^{-N/2} \left|\Omega^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(y_t - \Pi' x_t)' \Omega^{-1} (y_t - \Pi' x_t)\right]$$

Then the log-likelihood function is

$$l(\theta) = \sum_{t=1}^{T} \log f(y_t|x_t, \theta) = \frac{-TN}{2} \log(2\pi) + \frac{T}{2} \log(\Omega^{-1}) - \frac{1}{2} \sum_{t=1}^{T} (y_t - \Pi' x_t)' \Omega^{-1} (y_t - \Pi' x_t)$$

Choose $\hat{\Theta}$ to maximize $L(\Theta)$

$$\hat{\Pi}' = \left[\sum_{t=1}^{T} y_t x_t'\right] \left[\sum_{t=1}^{T} x_t x_t'\right]^{-1}$$

This is the OLS equation by equation.

Useful matrix calculation results:

1. Consider a quadratic form in A non-symmetric

$$\frac{\partial x' A x}{\partial a_{ij}} = x_i x_j$$

$$\frac{\partial x' A x}{\partial A} = x x'$$

2.

$$\frac{\partial \log |A|}{\partial A} = (A')^{-1}$$

Then we have

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{T}{2} \Omega' - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' = 0$$

Then

$$\hat{\Omega}' = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t'$$

4.3 Choice of Lag Length

Given $\hat{\Omega}$, value of $l(\hat{\theta})$

$$l(\hat{\Omega},\hat{\Pi}) = -\frac{TN}{2}\log(2\pi) + \frac{T}{2}\log\left|\hat{\Omega}^{-1}\right| - \frac{1}{2}\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t = -\frac{TN}{2}(12+\log(2\pi)) + \frac{T}{2}\log\left|\hat{\Omega}^{-1}\right|$$

where

$$\frac{1}{2}\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t = \frac{1}{2}tr\left(\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t\right) = \frac{1}{2}tr\left(\sum_{t=1}^{T}\hat{\Omega}^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right) = \frac{1}{2}tr\left(\hat{\Omega}^{-1}T\hat{\Omega}\right) = \frac{TN}{2}$$

where

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'$$

Suppose we want to test lag length $p_0 < p_1$. p_0 imposes $N^2(p_1-p_0)$ 0 restrictions.

$$2(l_1-l_0) = 2\left(\frac{T}{2}\log\left|\hat{\Omega}^{-1}\right| - \frac{T}{2}\log\left|\hat{\Omega}^{-1}\right|\right) = T\left(\log\left|\hat{\Omega}^{-1}\right| - \log\left|\hat{\Omega}^{-1}\right|\right) \sim \chi^2_N(p_1-p_0)$$

For small sample, Sim (1980) argues that

$$2(l_1-l_0) = (T-k)\left(\log\left|\hat{\Omega}^{-1}\right| - \log\left|\hat{\Omega}^{-1}\right|\right)$$

where $k = 1+Np_1$. The other ones are Akaike Information Criterion and Schwarz Information Criterion. They say choose the lag length that minimizes the $\log\left|\hat{\Omega}_p\right| + (pN^2+N)\frac{C(T)}{T}$ where $C(T)=2$ for AIC and $\log(T)$ for SIC.

4.4 Cambell (1991) and Hodrick (1992)

Let $z_t = [\log R_t, D_t/P_t, rb_t]'$ where $rb_t = i_t - \frac{\sum_{j=1}^{12} i_{t-j}}{12}$ (detrend interest rate, relative build rate), R_t is continuous compounded return, D_t is dividend rate and P_t is price.

Here $z(t)$ is de-meaned.

$$z_{t+1} = Az_t + u_{t+1}$$

$$(I-AL)z_{t+1} = u_{t+1} \implies z_{t+1} = (I-AL)^{-1}u_{t+1} = u_{t+1} + Au_t + A^2u_{t-1} + \cdots$$

u_{t+1} is serially uncorrelated and let $E[u_{t+1}u_{t+1}'] = V$ is the innovation covariance matrix. The unconditional variance of z_{t+1} is equal to

$$\begin{aligned} C(0) &= E[z_{t+1}z_{t+1}'] = E[(u_{t+1}+Au_t+A^2u_{t-1}+\cdots)(u_{t+1}+Au_t+A^2u_{t-1}+\cdots)'] \\ &= V + AVA' + A^2VA'^2 + \cdots \\ C(0) &= \sum_{j=0}^{\infty} A^jVA'^j \end{aligned}$$

$$E[z_{t+1}z_{t+1}'] = E[(Az_t+u_{t+1})(Az_t+u_{t+1})'] = AE[z_tz_t']A' + E[u_{t+1}u_{t+1}']$$
$$C(0) = AC(0)A' + V$$

Hamilton Proposition 10.4 states

$$vec(XYZ) = (Z'\otimes X)vec(Y)$$

where vec is a stack operator.

Then

$$vec(C(0)) = vec(AC(0)A') + vec(V) = (A\otimes A)vec(C(0)) + vec(V) = [I_{N^2} - A\otimes A]^{-1}vec(V)$$

$$C(1) = E[z_{t+1}z_t'] = E[(Az_t + u_{t+1})(z_t')] = AE[z_t z_t'] = AC(0)$$
$$C(2) = E[z_{t+2}z_t'] = E[(A^2 z - t + Au_{t+1} + u_{t+2})z_t'] = A^2 C(0)$$
$$C(j) = E[z_{t+j}z_t'] = A^j C(0)$$
$$C(-j) = C(j)'$$

Suppose we are interested in long horizon predictability. Let

$$\log R_{t+k,k} = \log R_{t+1} + \cdots + \log R_{t+k}$$

What is the variance of $\log R_{t+k,k}$. First, we will get the variance of $\sum_{j=1}^{k} z_{t+j}$

$$V_k = E[(z_{t+1} + z_{t+2} + \cdots + z_{t+k})(z_{t+1} + z_{t+2} + \cdots + z_{t+k})']$$

$$V_k = kC(0) + (k-1)(C(1) + C(-1)) + \cdots + (C(k-1) + C(-k+1) = kC(0) + \sum_{j=1}^{k-1}(k-j)(C(j) + C(j)')$$

(Note $V(\log R_{t+k,k}) = e_1' V_k e_1$ where $e_1 = (1,0,0)'$)

Fama-French looked at

$$\log R_{t+k,k} = \alpha_{k,1} + \beta_{k,1}\frac{D_t}{P_t} + u_{t+k,k}$$

we can use GMM with overlapping data.

$$\beta_{k,1} = Cov(\log R_{t+1} + \cdots + \log R_{t+k}, D_t/P_t)/Var(D_t/P_t)$$

But from VAR, all auto-covariances are determined. In particular, we can get

$$\beta_{k,1} = \frac{e_1'[C(1) + \cdots + C(k)]e_2}{e_2' C(0) e_2} = \frac{e_1'(A + A^2 + \cdots + A^k)C(0)e_2}{e_2' C(0) e_2}$$

This is implied slope coefficients. We can use delta method to get the variance.

The k-period variance ratio is

$$VR_k = \frac{Var(\log R_{t+1} + \cdots + \log R_{t+k}}{kVar(R_{t+1})} = \frac{e_1'[kC(0) + \sum_{j=1}^{k-1}(k-j)(C(j) + C(j)')]e_1}{ke_1' C(0) e_1}$$

R^2 from implied regression is the explained variance over the total variance that is

$$R_1^2(1) = \frac{\beta_{k,1}^2 e_2' C(0) e_2}{e_1' V_k e_1}$$

Explanatory Power of VAR at k horizon is

$$R_2^2(k) = 1 - \frac{\text{Innovation Variance}}{\text{Total Variance}}$$

requires the k-period innovation variance. $u_{t+1,1} = u_{t+1}$ at $k = 1$. This is innovation in z_{t+1}.

$$u_{t+2,2} = u_{t+2} + Au_{t+1} \text{ at } k = 2$$

$$u_{t+3,3} = u_{t+3} + Au_{t+2} + A^2 u_{t+1} \text{ at } k = 3$$

$$u_{t+k,k} = [I + AL + A^2L^2 + \cdots + A^{k-1}L^{k-1}]u_{t+k} = [(I - AL)^{-1}(I - A^k L^k)]u_{t+k}$$

The innovation variance is

$$\sum_{j=1}^{k}(I - A)^{-1}(I - A^j)V(I - A^j)'(I - A)'^{-1} = W_k$$

$$R_2^2(l) = 1 - \frac{e_2' W_k e_2}{e_1' V_k e_1}$$

Midterm March 7, 9-12 am, Uris 332, Closed Book and Notes

4.5 Impulse Response Functions

Univariate y_t, $E_t[y_{t+s}] - E_{t-1}[y_{t+s}]$ response to a shock $\varepsilon_t = 1$.

$$y_t = \alpha + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}, \theta_0 = 1$$

It's impulse response functions are the following

$$E_t[y_{t+1}] - E_{t-1}[y_{t+1}] = \theta_1$$
$$\vdots$$

Consider VAR, $y_t = \mu + \Phi y_{t-1} + \varepsilon_t$. Then

$$E[\varepsilon_t \varepsilon_t'] = \Omega, \text{full rank}$$

$$y_t = (I - \Phi L)^{-1}(\mu + \varepsilon_t) = (I - \Phi)^{-1}\mu + \sum_{j=0}^{\infty} \Phi^j \varepsilon_{t-j}$$

We are interested in impulse response function of $y_{k,t+j}$ to the shock of $\varepsilon_{k,t}$, that is

$$e_k' \Phi^j e_k$$

where e_k and e_k are indicator vectors. In other words, the following is equivalent

$$e_k' \Psi_j e_k$$

where $y_t = \mu + \sum_{j=0}^{\infty} \Psi \varepsilon_{t-j}$ $\varepsilon_{h,t} = 1$ with $\varepsilon_{j,t} = 0$ if $j \neq h$ makes no sense because it never happens in the world (you cannot really test this).

Ω is real symmetric positive definite so we can write

$$\Omega = ADA'$$

where A is lower triangular with 1's on the diagonal with positive entries off diagonal and zero elsewhere and D is a diagonal matrix. Now let's consider a process $u_t = A^{-1}\varepsilon_t$. Hence we have

$$E[u_t u_t'] = A^{-1}E[\varepsilon_t \varepsilon_t'](A^{-1})' = A^{-1}\Omega(A^{-1})' = A^{-1}ADA'(A^{-1})' = D$$

so u_t's are mutually uncorrelated. How let's consider

$$\begin{aligned} Au_t &= \varepsilon_t \\ u_{1t} &= \varepsilon_{1t} \\ u_{2t} &= \varepsilon_{2t} - a_{21}u_{1t} \\ \vdots &= \vdots \\ u_{jt} &= \varepsilon_{jt} - a_{j1}u_{t1t} - a_{j2}u_{2t} - \cdots - a_{j,j-1}u_{j-1,t} \end{aligned}$$

Because u_{jt} are uncorrelated, u_{jt} is the projection error of ε_{jt} onto $(u_{1t}, \cdots, u_{j-1,t})$ and a_{jk} are projection coefficients. Let x_t be $(y_t, y_{t-1}, \cdots)$. Then we have

$$\varepsilon_{1t} = y_{1t} - E[y_{1t}|x_{t-1}], \cdots, \varepsilon_{jt} = y_{jt} - E[y_{jt}|x_{t-1}]$$

The change in the projection

$$\frac{\partial \hat{E}[\varepsilon_{jt}|y_{1t}, x_{t-1}]}{\partial y_{1t}} = a_{j1}$$

For the vector we have

$$\frac{\partial \hat{E}[\varepsilon_t|y_{1t}, x_{t-1}]}{\partial y_{1t}} = a_1$$

Consequently,

$$\frac{\partial \hat{E}[y_{t+s}|y_{1t}, x_{t-1}]}{\partial y_{1t}} = \Psi_s a_1$$

This is the orthogonalized impulse response function. The issue is that the orthogonalization requires theory to make sense.

4.6 Variance Decompositions

What percent of forecast error variance is due to u_{jt}? We know that

$$y_{t+s} - \hat{y}_{t+s} = \varepsilon_{t+s} + \Psi_1 \varepsilon_{t+s-1} + \Psi_2 \varepsilon_{t+s-2} + \cdots + \varepsilon_{t+1}$$

$$MSE(y_{t+s}|t) = \Omega + \Psi_1 \Omega \Psi_1' + \cdots + \Psi_{s-1} \Omega \Psi_{s-1}$$

$$\Omega = ADA', D_{jj} = var(u_{jt})$$

$$\Omega = a_1 a_1' var(u_{1t}) + a_2 a_2' var(u_{2t}) + \cdots + a_m a_m' var(u_{mt})$$

$$MSE(y_{t+s}|t) = \sum_{j=1}^{m} var(u_{jt})[a_j a_j' + \Psi_1 a_j a_j' \Psi_1' + \cdots + \Psi_{s-1} a_j a_j' \Psi_{s-1}]$$

The contribution of u_{jt} is

$$var(u_{jt})[a_j a_j' + \Psi_1 a_j a_j' \Psi_1' + \cdots + \Psi_{s-1} a_j a_j' \Psi_{s-1}]$$

Since $MSE \to \Gamma_0$, the variance of y_t as $s \to \infty$, then it becomes the unconditional variance.

4.7 Models of Non-Stationarity Time Series

Hamilton Chapter 15 and Hayashi Chapter 9.

When we have stationary processes

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \psi_0 = 1$$

where $\sum_{j=0}^{\infty} |\psi_j| < \infty, \psi(z) = .0$ has roots outside the unit circle.

- $E[y_t] = \mu$
- $E[y_{t+s}|y_t y_{t-1}, \cdots] \to \mu$ as $s \to \infty$

However in general, economics and finance data are not stationary. We can take natural log. There are a few methods that attempt to solve the non-stationarity problem.

- Deterministic time trend

  $$y_t = \mu + \delta t + \psi(L)\varepsilon_t$$

  where $\psi(L)\varepsilon_t$ is as above. Here y_t is trend stationary. Campbell, Lettau, Malkiel, Xu (2001) JF argues that the aggregate idiosyncratic volatility of returns had a trend.

- Unit-Root Processes

  $$(1 - L)y_t = \delta + \psi(L)\varepsilon_t$$

  The last part should be stationary. We also assume $\psi(1) \neq 0$. y_t process is stationary after the first difference, e.g. log of GDP, log of price, log of exchange rate.

  $$\Delta \ln GDP_t = \text{rate of growth}$$

  $$\Delta \ln P_t = \text{rate o inflation change}$$

  $$\Delta \ln S_t = \text{change rate of appreciation rate}$$

  Why $\psi(1) \neq 0$? Suppose y_t is stationary, $y_t = \mu + \chi(L)\varepsilon_t$ is stationary. Then $(1 - L)y_t$ is also stationary and $(1-L)y_t = (1-L)\chi(L)\varepsilon_t = \psi(L)\varepsilon_t$ but $\psi(1) = 0$ in this case. It rules out starting with a stationary process.

  The prototypical unit root process is a random walk with drift, that is

  $$y_t = \delta + y_{t-1} + \varepsilon_t, \varepsilon_t \text{ i.i.d}$$

$$dy_t = \delta + \varepsilon_t$$

The unit root processes are integrated of order 1.

$$\frac{dy(t)}{dt} = x(t) \implies y(t) = \int x(t)dt$$

$$\Delta y_t = x_t$$

is

$$y_t = x_t + y_{t-1}$$
$$y_{t-1} = x_{t-1} + y_{t-2}$$
$$\vdots$$
$$y_t = \sum_{j=0}^{\infty} x_{t-j}$$

where $y - t$ is the sum over time of x_t.

By analogy, we get $I(2)$ are integrated of order 2.

$$(1-L)^2 y_t = k + \psi(L)\varepsilon_t$$

$ARMA(p,q)$ was stationary $AR(p)$, $MA(q)$. $ARIMA(p,d,q)$ so difference d times and then $AR(p)$ and $MA(q)$ processes.

$$\phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t$$

Typically (1, 1, 1) is enough.

4.7.1 Compare Forecasts

If Y_t is the level of GDP, $y_t = \ln(Y_t)$ then

$$\Delta y_t = \text{growth rate of GDP}$$

This change can be population, labor force participation, investment and technology change. They are usually stationary.

Trend Stationary

$$y_t = \alpha + \delta t + \psi(L)\varepsilon_t$$
$$y_{t+s} = \alpha + \delta(t+s) + \psi(L)\varepsilon_{t+s}$$
$$\hat{y}_{t+s,t} = E[y_{t+s}|y_t, \cdots] = \alpha + \delta(t+s) + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \cdots$$
$$E[\hat{y}_{t+s,t} - \alpha - \delta(t+s)] \to 0$$

as ψ_j dies out

The forecast errors

$$y_{t+s} - \hat{y}_{t+s,t} = \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \cdots + \psi_{s-1}\varepsilon_{t+1}$$

The MSE of the Forecast is $\sigma^2(1 + \psi_1^2 + \cdots + \psi_{s-1}^2)$. As $s \to \infty$, MSE goes to unconditional variance of $\psi(L)\varepsilon_t$.

Unit Root

$$\Delta y_t = \delta + \psi(L)\varepsilon_t$$

$$\begin{aligned} y_{t+s} &= \Delta y_{t+s} + \Delta y_{t+s-1} + \cdots \Delta y_{t+1} + y_t \\ &= (\delta + \psi(L)\varepsilon_{t+s}) + (\delta + \psi(L) + \varepsilon_{t+s-1}) + \cdots + (\delta + \psi(L)\varepsilon_{t+1}) + y_t \end{aligned}$$

$$\hat{y}_{t+s,t} \to s\delta + y_*$$

as $s \to \infty$

The forecast errors

$$\begin{aligned}
y_{t+s} - \hat{y}_{t+s,t} &= \Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+1} + y_t - [\Delta \hat{y}_{t+s,t} + \cdots + \Delta \hat{y}_{t+1,t} + \hat{y}_t] \\
&= (\varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \cdots + \psi_{s-1}\varepsilon_{t+1}) \\
&+ (\varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \cdots + \psi_{s-2}\varepsilon_{t+1}) \\
&+ \vdots \\
&= \varepsilon_{t+1} \\
&= \varepsilon_{t+s} + (1+\psi_1)\varepsilon_{t+s-1} + (1+\psi_1+\psi_2)\varepsilon_{t+s-2} + \cdots + (1+\psi_1+\psi_2+\cdots+\psi_{s-1})\varepsilon_{t+1} \\
MSE &= \sigma^2[1 + (1+\psi_1)^2 + \cdots + (1+\psi_1+\psi_2+\cdots+\psi_{s-1})^2
\end{aligned}$$

4.8 Hodrick-Prescott Filter

Let $y_t = \log(GDP) = g_t + c_t$ where g_t is a smooth trend (Δg_t is stationary) and c_t is a cyclical component. We want to minimize the cyclical components subject to g_t not varying very much.

$$\min_{\{g_t\}_{t=1}^T} \left\{ \sum_{t=1}^T (y_t - g_t)^2 + \lambda \sum_{t=1}^T ((g_t - g_{t-1}) - (g_{t-1} - g_{t-2}))^2 \right\}$$

where quarterly data uses $\lambda = 600$ (A particular unobservable components model).

4.9 Special Cases

Random walk with drift

$$\hat{y}_{t+s,t} = s\delta + y_t + \varepsilon_{t+s}$$

log series is expected to grow at the rate of δ from wherever it is y_t.

ARIMA(0, 1, 1):

$$\Delta y_t = \delta + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\hat{y}_{t+1,t} = \delta + y_t + \varepsilon_{t+1}$$

$$y_{t+1} - \hat{y}_{t+1,t} = \theta \varepsilon_t$$

$$\varepsilon_t = y_t - \hat{y}_{t,t-1}, \text{for } \delta = 0$$

$$\hat{y}_{t+1,t} = y_t + \theta(y_t - \hat{y}_{t,t-1}) = (1+\theta)y_t - \theta \hat{y}_{t,t-1}$$

For $|\theta| < 1$

$$(1 + \theta L)\hat{y}_{t+1,t} = (1+\theta)y_t$$

$$\hat{y}_{t+1,t}. = \frac{(1+\theta)y_t}{1 - (-\theta L)} = (1+\theta)\sum_{j=0}^{\infty}(-\theta)^j y_{t-j}$$

This is exponential smoothing. If $\theta < 0$. the right hand side is how people formed expectations in 1960s. Friedman (1957) says it permanent increases. Muth (1961) says exponential smoothing is only rational if series is $(0, 1, 1)$.

4.10 Beveridge-Nelson Decomposition

Every unit-root process can be decomposed into a random walk with drift plus a zero-mean stationary component.

$$(1 - L)y_t = \mu + a(L)\varepsilon_t$$

where roots $a(z)$ are outside the unit circle.

$$y_t = z_t + c_t$$

where z_t is the random walk with drift and c_t is the stationary part.

The claim is that

$$z_t = \mu + z_{t-1} + a(1)\varepsilon_t$$

$$c_t = a^*(L)\varepsilon_t, a_j^* = -\sum_{k=j+1}^{\infty} a_k$$

Proof. Proof by construction.

$$(1-L)y_t = (1-L)z_t + (1-L)c_t$$

$$(1-L)z_t = \mu + a(1)\varepsilon_t$$

$$(1-L)c_t = (1-L)a^*(L)\varepsilon_t$$

$$a(1) = a_0 + a_1 + a_2 + a_3 + \cdots$$

$$(1-L)a_0^* = -a_1 - a_2 - a_3 - \cdots + a_1L + a_2L + \cdots$$

$$(1-L)a_1^*L = -a_2L - a_3L - \cdots + a_2L^2 + a_3L^2 + \cdots$$

$$\vdots$$

$$(1-L)y_t = \mu + (a(1) + (1-L)a^*(L))\varepsilon_t = \mu + a(L)\varepsilon_t$$

□

4.11 Fractional Integration

$ARIMA(p,d,q)$ implies $(1-L)^d y_t = \psi(L)\varepsilon_t$ for MA infinity representation. The impulse response function decays geometrically.

$$(1-\rho L)y_t = \varepsilon_t \implies y_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3\varepsilon_{t-3}$$

Granger, Jayeux (1980) and Hosking (1981) considers $[(1-L)^d]^{-1}$ exists for $d < \frac{1}{2}$

$$y_t = (1-L)^{-d}(\psi(L))\varepsilon_t$$

$$f(z) = (1-z)^{-d}$$

$$\frac{df}{dz} = d(1-z)^{-(d+1)}$$

$$\frac{d^2f}{dz^2} = (d+1)d(1-z)^{-(d+z)}$$

$$\vdots$$

Power series expansion of $f(z)$ around $z = 0$

$$f(z) = f(0) + \left.\frac{df}{dz}\right|_{z=0} z + \frac{1}{2!}\left.\frac{d^2f}{dz^2}\right|_{z=0} z^2 + \cdots + \cdots$$

$$(1-z)^{-d} = 1 + dz + \frac{1}{2}(d+1)dz^2 + \frac{1}{3!}(d+2)(d+1)dz^3 + \cdots$$

$$(1-L)^{-d} = \sum_{j=0}^{\infty} h_j L^j$$

where $h_0 = 1$ and $h_j = \frac{1}{j!}(d+j-1)(d+j-2)\cdots(d+1)d$ Therefore

$$y_t = (1-L)^{-d}\varepsilon_t = h_0\varepsilon_t + h_1\varepsilon_{t-1} + h_2\varepsilon_{t-2} + \cdots$$

Infinite order MA with particular impulse response function decays slowly than the geometric decay. (long memory; Bollerslev GARCH)

4.11.1 GARCH

$$y_t = \mu + \varepsilon_t$$

where

$$\varepsilon_t = N(0, h_t)$$

and

$$h_t = \omega + \beta h_{t-1} + \alpha \varepsilon_{t-1}^2 \text{ conditional variance processes}$$

$$h_t = E[\varepsilon_t^2]$$

$$E[h_t] = \omega + \beta E[h_{t-1}] + \alpha E[\varepsilon_{t-1}^2]$$

$$V = E[h_t] = E[h_{t-1}] = E[\varepsilon_{t-1}^2]$$

$$V = \frac{\omega}{1 - \alpha - \beta}, \alpha + \beta < 1$$

By applying the factional integration into GARCH, we call it FGARCH.

4.12 Testing For Unit-Root

Section 15.4 from Hamilton gives a good discussion about this topic. Suppose

$$y_t = y_{t-1} + \varepsilon_t$$

is the truth

$$y_t = \rho y_{t-1} + \varepsilon_t$$

set $\rho = .9999$ with 10,000 observation you won't be able to reject $\rho = .99999$ v.s. 1

Consider

$$y_t = \rho y_{t-1} + u_t$$

where u_t is i.i.d $N(0, \sigma^2)$ Estimate $\hat{\rho}$ with OLS

$$\hat{\rho} = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} = \frac{\sum_{t=1}^T y_{t-1}(\rho y_{t-1} + u_t)}{\sum_{t=1}^T y_{t-1}^2} = \rho + \frac{\sum_{t=1}^T y_{t-1} u_t}{\sum_{t=1}^T y_{t-1}^2}$$

when y_t is stationary

$$\sqrt{T}(\hat{\rho} - \rho) = \frac{1/\sqrt{T} \sum_{t=1}^T y_{t-1} u_t}{1/T \sum_{t=1}^T y_{t-1}^2}$$

where $\sqrt{T}(\hat{\rho} - \rho) \to N(0, \Omega)$ and $\Omega = Q^{-1} S Q^{-1}, Q = E[y_{t-1}^2]$ and $S = E[y_{t-1}^2, u_t^2] = Q\sigma^2$ with homoskedasticity.

$$\Omega = Q^{-1} Q \sigma^2 Q^{-1} = \sigma^2 Q^{-1}$$

$$Q = E[y_{t-1}^2] = \frac{\sigma^2}{1 - \rho^2}$$

$$\Omega = \frac{\sigma^2}{\sigma^2/(1 - \rho^2)} = (1 - \rho^2)$$

$$\sqrt{T}(\hat{\rho} - \rho) \to N(0, 1 - \rho^2)$$

Notice if $\rho = 1$, we would have $N(0, 0)$. It is impossible. The law of large numbers and convergence only work for $|\rho| < 1$.

Suppose rather by scaling by the root of T, let's scale by T.

$$T(\hat{\rho} - \rho) = \frac{1/T \sum_{t=1}^T y_{t-1} u_t}{1/T^2 \sum_{t=1}^T y_{t-1}^2}$$

If y_t is random walk, $\rho = 1$, let $y_0 = 0$

$$y_t = u_t + u_{t-1} + \cdots + u_1$$

$$y_t \sim N(0, t\sigma^2) \tag{2}$$

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2$$

$$y_{t-1}u_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - u_t^2)$$

Therefore,

$$\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}(y_T^2 - y_0^2) - 0.5\sum_{t=1}^{T} u_t^2 = 0.5y_T^2 - 0.5\sum_{t=1}^{T} u_t^2$$

$$1/T\sum_{t=1}^{T} y_{t-1}u_t = 0.5/Ty_T^2 - 0.5/T\sum_{t=1}^{T} u_t^2$$

Divide by σ^2

$$0.5/(\sigma^2 T)y_T^2 - 0.5/(\sigma^2 T)\sum_{t=1}^{T} u_t^2$$

This is equal to

$$0.5(y_T/(\sigma\sqrt{T}))^2 - 0.5/(\sigma^2 T)\sum_{t=1}^{T} u_t^2 = 0.5\chi^2(1) - 0.5 = 0.5(\chi^2(1) - 1)$$

The denominator is

$$E\left[\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2\right] = \frac{1}{T^2}\sigma^2\sum_{t=1}^{T}(\text{equation } 2 - 1) = \frac{\sigma^2(T-1)T}{T^2 2}$$

by functional central limit theory

You can demonstrate that $T(\hat{\rho} - 1) < 0$ 68% of the time, even though $\rho = 1$. Hayashi has Dikey-Fuller discussion on page 487 and table B5 on page 762.

$$y_t = y_{t-1} + \varepsilon_t$$

$$y_t = \rho y_{t-1} + \varepsilon_t$$

Calculate $T(\hat{\rho} - 1)$ for $T = 100$. The probability $T(\hat{\rho} - 1) < 131$ is 95% and $T(\hat{\rho} - 1) < -7.9$ is 5%. Reject $\rho = 1$ if $T(\hat{\rho} - 1) < -7.9$ at 5% critical value. $\hat{\rho} - 1 = \frac{1}{100}(-7.9)$. $\hat{\rho} = 1 - 0.079 = .921$. if $\hat{\rho} < .92$, you can reject $H_0 : \rho = 1$.

4.13 Cointegration

$$y_t(m \times 1)$$

each y_{it} is $I(1), i = 1, \cdots, m$. $a'y_t$ where a is $m \times 1$ vector of constants is stationary.

4.13.1 Purchasing Power Parity

$(\$/£) = S_t$. $P_t^{\$}$ is the dollar price level and $P_t^{£}$ is pound price level. Internal purchasing power is the

$$\frac{1}{P_t^{\$}} = \frac{\text{Goods}}{\$}$$

and the external purchasing power in the UK

$$\frac{1}{S_t} = \frac{£}{\$}, \frac{1}{P_t^{£}} = \frac{\text{Goods}}{£}$$

Then

$$\frac{1}{P_t^{\$}} = \frac{1}{S_t}\frac{1}{P_t^{£}}$$

Take logs

$$S_t^{\$/£} = P_t^{\$} - P_t^{£}$$
$$S_t \neq S_t^{PPP}$$
$$S_t - P_t^{\$} + P_t^{£} = \text{deviations from PPP}$$

where S_t is $I(1)$, $P_t^{\$}$ is $I(1)$ and $P_t^{£}$ is $I(1)$. Here $a' = (1, -1, 1)$ and $a'\begin{pmatrix} S_t \\ P_t^{\$} \\ P_t^{£} \end{pmatrix}$ = stationary process cointegration

4.14 Price-Dividend Ratio and Campbell-Shiller Decomposition

Let r_{t+1} = rate of return on a stock, p_t = log price of stock, d_t = log dividend.

$$\exp(r_{t+1}) = \frac{P_{t+1} + D_{t+1}}{P_t}$$

Factor out D_{t+1} and Divide numerator and denominator by D_t then

$$\exp(r_{t+1}) = \frac{\left[\frac{P_{t+1}}{D_{t+1}} + 1\right]\frac{D_{t+1}}{D_t}}{\frac{P_t}{D_t}}$$

Take logs

$$r_{t+1} = \log[\exp(p_{t+1} - d_{t+1}) + 1] + \Delta d_{t+1} - (p_t - d_t)$$
$$r_{t+1} = \log[1 + \exp(\overline{p-d})] + \frac{\exp(\overline{p-d})}{1 + \exp(\overline{p-d})}(p_{t+1} - d_{t+1} - \overline{p-d}) + \Delta d_{t+1}$$
$$r_{t+1} = \kappa + \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t)(\kappa \text{ is some constant term})$$

Holds ex post and ex ante. Take E_t of both side.

$$(p_t - d_t)(1 - \rho L^{-1}) = \kappa + \Delta d_{t+1} - r_{t+1}$$

$$p_t - d_t = \frac{\kappa}{1-\rho} + E_t \sum_{j=1}^{\infty} \rho^{j-1}(\Delta d_{t+j} - r_{t+j})(\text{Campbell-Shiller Decomposition})$$

Here r_{t+j} is stationary, d_{t+1} is $I(1)$, Δd_{t+1} is stationary, p_{t+1} is $I(1)$. Lastly, $(p_t - d_t)$ is stationary.

The CS decomposition is driven by the similarity of the dividend to earning's ratio $\frac{D_t}{EA_t}$ where $d_t - ea_t$ is stationary.

4.14.1 Lettau-Ludigson: log Consumption Wealth Ratio

$$cay_t = c_t - \alpha w_t$$

4.14.2 Hamilton's Canonical Example

2y's,

$$y_{1t} = \gamma y_{2t} + u_{1t}$$

$$y_{2t} = y_{2t-1} + u_{2t}$$

where u_{1t} and u_{2t} are serially uncorrelated white noise.

$$\Delta y_{2t} = u_{2t}$$

so y_{2t} is $I(1)$ and y_{2t} is ARIMA(0, 1, 0).

$$\Delta y_{1t} = \gamma \Delta y_{2t} + u_{1t} - u_{1t-1} = \gamma u_{2t} + u_{1t} - u_{1t-1} = r_t + \theta r_{t-1}$$

$\Delta y_{1t} = v_t + \theta v_{t-1}$ is $I(1)$ and y_{1t} is ARIMA(0, 1, 1).

$$y_{1t} - \gamma y_{2t} = u_{1t}$$

so this is stationary. Hence y_{1t} and y_{2t} are cointegrated with vector $(1, -\gamma)$.

For VAR representation, we need $\varepsilon_{1t} + \varepsilon_{2t}$ as forecast errors relative to Φ_{t-1}.

$$\varepsilon_{1t} = \gamma \varepsilon_{2t} + u_{1t}$$

$$\varepsilon_{2t} = u_{2t}$$

$$E_{t-1}(y_{1t}) = \gamma E_{t-1}(y_{2t})$$

$$y_{1t} - E_{t-1}(y_{1t}) = \varepsilon_{1t} = \gamma(y_{2t} - E_{t-1}(y_{2t})) + u_{1t} = \gamma \varepsilon_{2t} + u_{1t}$$

Hence

$$u_{1t} = \varepsilon_{1t} - \gamma \varepsilon_{2t}$$

Postulate stationary VAR in Δy_{1t} and Δy_{2t}.

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \Psi(L) \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Can we invert $\Psi(L)$ to get finite order VAR

$$\psi(L) = \begin{pmatrix} (1-L) & \gamma L \\ 0 & 1 \end{pmatrix}$$

$$\begin{aligned} \Delta y_{1t} &= \gamma \Delta y_{2t} + \Delta u_{1t} = \gamma u_{2t} + u_{1t} - u_{1t-1} \\ &= \gamma \varepsilon_{2t} + u_{1t} - (\varepsilon_{1t} - \gamma \varepsilon_{2t}) \\ &= \varepsilon_{1t} - \varepsilon_{1t-1} + \gamma \varepsilon_{2t} \\ &= (1-L)\varepsilon_{1t} + \varepsilon_{2t-1} \end{aligned}$$

We know that $\Psi(z)$ has a root at 1 so $|\Psi(1)| = 0$ so $\Psi(z)^{-1}$ does not exist. We have

$$\Delta y_{1t} = \gamma \Delta y_{2t} - \Delta u_{1t} = \Gamma u_{2t} + u_{1t} - u_{1t-1}$$

$$u_{1t-1} = y_{1t-1} - \gamma y_{2t-1}$$

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \gamma u_{2t} + u_{1t} \\ u_{2t} \end{pmatrix}$$

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Cointegrated VAR lapped cointegrated variable on the right hand side. This is error correction representation.

4.15 Normalizations

If $y_t(m \times 1)$ and each y_{it} is $I(1)$ and $a'y_t$ is stationary. "a" (cointegration factor) is not unique and for scalar b, ba also implies stationary process and $a_{11} = 1$ is a appropriate normalization. There may be $h < m$ unique cointegrating vectors. We can stack them in $A(m \times h)$ where

$$A' = \begin{pmatrix} a_1' \\ \vdots \\ a_h' \end{pmatrix}$$

Δy_t is stationary and $\delta = E[\Delta y_t]$ define $u_t = \Delta y_t - \delta$. Write the Wold Decomposition of u_t as

$$u_t = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots = \Psi(L)\varepsilon_t$$

$$E[\varepsilon_t \varepsilon_{t-s}'] = \begin{cases} \Omega & s = 0 \\ 0 & s \neq 0 \end{cases}$$

$$\Psi(1) = I_m + \Psi_1 + \Psi_2 + \cdots$$

Claim: If $A'y_t$ is stationary, then the necessary conditions are

$$A'\Psi(1) = 0$$

$$A'\delta = 0$$

Proof.

$$\Delta y_t = \delta + \Psi(L)\varepsilon_t$$

a vector MA representation. Iterate into the path and get

$$y_t = y_0 + \delta t + (u_t + u_{t-1} + \cdots + u_1)$$

Do the Beveridge-Nelson decomposition, we say

$$\Psi(L) = \Psi(1) = (1 - L)\alpha(L)$$

where $\alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j, \alpha_j = (\Psi_{j+1} + \Psi_{j+2} + \cdots)$

$$u_t = \Psi(L)\varepsilon_t = \Psi(1)\varepsilon_t + \alpha(L)(\varepsilon_t - \varepsilon_{t-1})$$

Define $\eta_t = \alpha(L)\varepsilon_t$ stationary substitute for u_t's

$$y_t = y_0 + \delta t + [\Psi(1)\varepsilon_t + (\eta_t - \eta_{t-1}) + \Psi(1)\varepsilon_{t-1} + (\eta_{t-1} - \eta_{t-2}) + \cdots + \Psi(1)\varepsilon_1 + \eta_1 - \eta_0]$$

$$y_t = y_0 + \delta t + \Psi(1)[\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1] + \eta_t - \eta_0$$

This is the multivariate Beveridge-Nelson.

$$A'y_t = A'y_0 + A'\delta t + A'\psi(1)[\varepsilon_t + \cdots + \varepsilon_1] + A'\eta_t - A'\eta_0$$

so $A'\delta = 0$ and $A'\Psi(1) = 0$ for stationarity.

$$\Psi(z) = \begin{pmatrix} 1 - z & \gamma_t \\ 0 & 1 \end{pmatrix}$$

$$\Psi(1) = \begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix}$$

$$a' = (1, -\gamma)$$

$$a'\Psi(1) = (1, -\gamma)\begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix} = 0$$

□

4.16 Triangular Representation

$$A' = \begin{pmatrix} a_1' \\ \vdots \\ a_h' \end{pmatrix}$$

where $A'y_t$ is stationary vectors, $A'\delta = 0$ and $A'\Psi(1) = 0$.

$$A' = \begin{pmatrix} 1 & 0 & \cdots & -\gamma_{1,h+1} & -\gamma_{1,h+2} & \cdots & -\gamma_{1,m} \\ 0 & 1 & \cdots & -\gamma_{2,h+1} & -\gamma_{2,h+2} & \cdots & -\gamma_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & \cdots & 1 & -\gamma_{h,h+1} & -\gamma_{h,h+2} & \cdots & -\gamma_{h,m} \end{pmatrix} = [I, -\Gamma_{h\times(m-h)}]$$

$z_t = A'y_t$ and $E[z_t] = \mu_1^*$ Partition

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$$

Demeaned z_t, $z_t^* = z_t - \mu_1^*$.

$$z_t^* + \mu_1^* = A'y_t = [I_n, -\Gamma] \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$$

$$y_{1t} = \Gamma y_{2t} + \mu_1^* + z_t^*$$

$$\Delta y_{2t} = \delta_2 + u_{2t}$$

where $u_{2t} = E[\Delta y_{2t}]$ is serially correlated. Write the stationary components as a Wold decomposition.

$$\begin{pmatrix} z_t^* \\ u_{2t} \end{pmatrix} = \sum_{j=0}^{\infty} \begin{pmatrix} H_s \\ J_s \end{pmatrix} \varepsilon_{t-s}$$

where ε is m by 1, H is h by m and J is g by m. With Beveridge-Nelson Decomposition, we have

$$y_{2t} = y_{2s} + \delta_2 t + J(1)[\varepsilon_1 + \cdots + \varepsilon_t] + \eta_{2t} - \eta_{2s}$$

We have

$$y_{2t} = \tilde{u}_2 + \delta_2 t + \zeta_{2t} + \eta_{2t}$$

where $\tilde{u}_2 = y_{2s} - \eta_{2s}$, ζ_{2t} is random walk and η_{2t} is stationary.

$$y_{1t} = \Gamma y_{2t} + u_1^* + z_t^*$$

$$y_{1t} = u_1^* + \Gamma(\tilde{u}_2 + \delta_2 + \zeta_{2t} + \eta_{2t}) + z_t^*$$

$$y_{1t} - \tilde{u}_1^* + \Gamma(\delta_2 + \zeta_{2t}) + \tilde{\eta}_{1t}$$

where $\tilde{u}_1 = \mu_1^* + \Gamma\tilde{u}_2, \tilde{\eta}_{1t} = z_t^* + \Gamma\eta_{2t}$ This is the Stock-Watson Common Trends Representation of y_t series. y_t is linear-combination of g deterministic trends $\delta_2 t$ and g common random walks ζ_{2t}. and stationary components

$$\begin{pmatrix} \tilde{u}_1 \\ \tilde{u}_2 \end{pmatrix} + \begin{pmatrix} \tilde{\eta}_{1t} \\ \tilde{\eta}_{2t} \end{pmatrix}$$

4.17 Error Correlation VAR

y_t as p-th order non-stationary VAR.

$$y_t = d + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t$$

$$\Phi(L)y_t = d + \varepsilon_t, \Phi(L) = I - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$$

$$\Delta y_t = \delta + \Psi(L)$$

$$(\Delta - L)\Phi(L)y_t = \Phi(1)\delta + \Phi(L)\Psi(L)\varepsilon_t$$

$$(1-L)(\alpha+\varepsilon_t)=\Phi(1)\delta+\Phi(L)\Psi(L)\varepsilon_t$$
$$(1-L)\alpha=0, \Phi(1)\delta=0 \text{ is required}$$
$$(1-L)I_m=\Phi(L)\Psi(L) \text{ identical. polynomial in lag operator.}$$
$$(1-z)I_m=\Phi(z)\Psi(z), z=1 \implies \Phi(1)\Psi(1)=0$$

For any row $\Phi(1)$ denote Π',

$$\Pi'\Psi(1)=0, \Pi'\delta=0$$

determines the cointegration vector.

$$\Pi=Ab, \Pi'=b'A', \forall \text{ rows of } \Phi(1), \Phi(1)=BA'$$

. Hence $\Phi(1)$ is singular.

$$|I_m-\Phi_1 z-\Phi_2 z^2-\cdots\Phi_p z^p|=0$$

at $z=1$. There is at least unit root.

$$y_t=\alpha+\Phi_1.y_{t-1}+\cdots+\Phi_p y_{t-p}+\varepsilon_t$$

$$y_t=\rho_1\Delta y_{t-1}+\rho_2\Delta y_{t-2}+\cdots+\rho_{p-1}\Delta y_{t-p-1}+\alpha+\rho+\varepsilon_t$$

where

$$\rho=\Phi_1+\cdots+\Phi_p$$
$$\rho_s=-[\Phi_{s+1}+\cdots+\Phi_p], s=1,\cdots,p-1$$

Subtract y_{t-1}

$$y_t-y_{t-1}=\Delta y_t=\rho_1\Delta y_{t-1}+\cdots+\rho_{p-1}\Delta y_{t-p+1}+\alpha.+(\rho-I)y_{t-1}+\varepsilon_t$$
$$[\rho-I]=-[I-\Phi_1-\cdots-\Phi_p]=-\Phi(1)=-BA'$$
$$\Delta y_t=\alpha\Delta y_{t-1}+\cdots+\delta_{p-1}\Delta y_{t-p+1}-BA'y_{t-1}\varepsilon_t$$

4.18

PPP theory says

$$z_t=S_t-P_t^{\$}+P_t^{£}$$

is stationary where $a=(1,-1,1)$. Then we can use DF to test unit root for each series and z_t. Then z_t is stationary and $S_t, P_t^{\$}, P_t^{£}$ are cointegrated. In Hamilton, with lira and dollar exchange rate, each was $I(1)$ but Z_t could not reject unit root. Normalize $a_n=1$ and estimate $(n-1)$ cointegrating parameters.

$$y_{1t}=\gamma_2 y_{2t}+\gamma_3 y_{3t}+\cdots+\gamma_m y_{mt}+\varepsilon_{1t}$$

Minimize the sum of squared residual which is the second moment of z_t if there is cointegration

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{1t}^2\to E[z_t^2]$$

if cointegration; otherwise, their $\frac{1}{T}SSR$ diverges and $\frac{1}{T^2}SSR$ converges to Brownian motion.