

Apuntes Econometría II

1 Series de tiempo univariadas

1.1 Conceptos y definiciones

- **Diferencia entre momentos condicionales y momentos incondicionales:** **Momentos condicionales** son condicionales a un conjunto de información con el cuál estamos intentando hacer alguna evaluación en particular. **Momentos incondicionales** responde a una evaluación que no depende de una condición inicial (ni del tiempo).
- **Autocovarianza:** Es uno de los segundos momentos. \mathcal{E}_t sería la esperanza condicional en la información hasta t , mientras que \mathcal{E} sería la esperanza incondicional. Para las autocorrelaciones, tendremos:

$$\gamma_{0t} = \mathcal{E}(Y_t - \mu_t)^2, \quad \gamma_{it} = \mathcal{E}(Y_t - \mu_t)(Y_{t-i} - \mu_{t-i})$$

Donde μ_t corresponde al promedio de la serie en el momento t . Si la serie es estacionaria, entonces sabemos que $\mu_t = \mu$. Note que autocovarianza vendría siendo la covarianza de la serie respecto de sí misma rezagada en i períodos.

- **Autocorrelación:** Viene de correlación de la serie con sí misma rezagada i períodos. Es una manera natural de medir persistencia de una serie, pues es una variable libre de escala, pues se normaliza al dividir por la varianza de la serie. Con esto, podemos medir cercanía y movimiento de las series sin depender de la escala de las mismas.

$$\rho_{it} = \gamma_{it} / \gamma_{0t}$$

Por definición, la autocorrelación estará entre $0 \leq |\rho_{it}| \leq 1$. Esto implica que las autocorrelaciones pueden ser positivas o negativas. Así, ellas nos entregan información de proximidad y de dirección.

- **Proceso débilmente estacionario** (estacionario en covarianza): Se da cuando los dos primeros momentos de la serie existen y no dependen de t . Esto es, la esperanza y las covarianzas del proceso no deben depender del momento del tiempo en el que se miren. Luego, no debe haber nada que haya cambiado en la estructura de la serie. Ahora, sí puede ocurrir que $\gamma_i \neq \gamma_j$, $j \neq i$
- **Estacionariedad Estricta:** Concepto más fuerte; función de distribución conjunta de la serie Y en el período $t, t-1, t-2, \dots$ no depende de t . En otras palabras, nos indica que todos los momentos que pueden derivarse de esa serie, no dependen de t . Luego, al menos en las series que veremos en el curso, $FE \Rightarrow DE$, pero no al revés.
- **Ergodicidad:** Antes que nada, se define siempre respecto un momento.
 - Proceso DE es ergódico a su media si

$$T^{-1} \sum y_t \xrightarrow{p} E(y_t)$$

- Proceso DE es ergódico para sus segundos momentos si

$$(T-j)^{-1} \sum (y_t - \mu)(y_{t-j} - \mu) \xrightarrow{p} \gamma_j$$

Basicamente, es revisar que el cálculo del momento muestral converge en probabilidad al momento poblacional.

- **Ruido Blanco:** Un proceso es un ruido blanco si es que sus autocovarianzas son iguales a 0 i.e. $\gamma_j = 0, \forall j \geq 1$. En palabras, no hay nada en el pasado de la serie que nos ayude a predecir cuál será el futuro de la misma. Típicamente se pide que los errores de estimación tengan una estructura de ruido blanco, ya que si no, hay una estructura sistemática no modelada en la práctica, por lo que no estamos utilizando eficientemente la información que se tiene.
- **Innovación:** Proceso que, además de ser ruido blanco, es ortogonal a todo el set de información disponible. Esto es, que no hay nada que se conozca que pueda ayudar a predecir el comportamiento de la serie. Es un concepto mucho más fuerte que ruido blanco.

1.2 Tipos de procesos: MA(1)

Moving Average. Se describe de la siguiente manera:

$$y_t = \mu + u_t + \theta u_{t-1}, \quad u_t \sim iid(0, \sigma^2)$$

Donde θ es el parámetro que define al proceso. Por lo tanto, tendremos que $E(u_t, u_s) = 0, \forall t \neq s$. No hay correlación entre los errores de cada período. La dependencia temporal es entre la serie y_t , donde tenemos que u_t estará tanto en y_t como en y_{t+1} . Ahora, describiremos las propiedades del proceso:

- $E(y_t) = E(\mu + u_t + \theta u_{t-1}) = \mu$
- $\gamma_0 \equiv E(y_t - \mu)^2 = E(u_t + \theta u_{t-1})^2 = (1 + \theta^2)\sigma^2$
- $\gamma_1 = E(y_t - \mu)(y_{t-1} - \mu) = \theta\sigma^2$
- $\gamma_j = E(y_t - \mu)(y_{t-j} - \mu) = 0, \forall j \geq 2$
- $\rho_1 = \frac{\theta}{1+\theta^2}$
- $\rho_j = \gamma_j/\gamma_0 = 0, \forall j \geq 2$

Note que el signo de la autocovarianza y la autocorrelación dependerán netamente del signo de θ . Ahora, veamos las funciones de impulso respuesta. Para computarlas, se podría tomar la derivada de la serie respecto del ruido blanco. Luego, nos interesa encontrar $\frac{\partial y_{t+s}}{\partial u_t}$. Por ejemplo, tenemos que $\frac{\partial y_t}{\partial u_t} = 1$. Con esto, estamos calculando cuál es el impulso que estamos dando en el período presente. Nos gustaría ver cuál es la respuesta a ese impulso que tiene la serie, viendo como cambia la trayectoria temporal de y_t en cada uno de los períodos siguientes. Así, tendremos que para este caso, la función de impulso respuesta tiene la siguiente forma:

$$\frac{\partial y_{t+j}}{\partial u_t} = \begin{cases} \theta^j & j = 0, 1 \\ 0 & j \geq 2 \end{cases}$$

Para este caso, sabemos que la memoria del proceso es finita y además exactamente igual al orden del proceso que estamos estudiando. Por tanto, todo lo que hemos derivado no depende del valor numérico que pueda tener el parámetro θ

1.3 Tipos de procesos: $MA(q)$

Para este caso, es simplemente la generalización del proceso anterior. Es decir, ahora dependeremos de hasta q rezagos de ruidos blancos.

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}$$

Note que q vendría siendo el máximo de rezagos que se encuentra en la serie. Así, pueden haber algunos parámetros que tengan valor igual a 0.

A continuación, mostramos sus momentos

- $\mathcal{E}(y_t) = \mu$
- $\gamma_0 = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)$
- $\gamma_j = \begin{cases} \sigma^2(\theta_j + \theta_{j+1}\theta_1 + \cdots + \theta_q\theta_{q-j}) & j = 1, \dots, q \\ 0 & j > q \end{cases}$

Las IRF de este caso será análoga al proceso $MA(1)$, por lo que tendremos que

$$\frac{\partial y_{t+j}}{\partial u_t} = \theta_j$$

Ahora, pensemos en algo diferente. Supongamos que $q \rightarrow \infty$. Así, podríamos representar el proceso $MA(\infty)$ como:

$$y_t = \mu + \sum_{j=0}^{\infty} \theta_j u_{t-j} \quad (\theta_0 = 1)$$

Y en este contexto, cabe preguntarse por condiciones que garanticen estacionariedad débil del proceso. Y en ese sentido, tendremos que un $MA(\infty)$ será estacionario si se cumple alguna de las siguientes condiciones:

- **Sumabilidad al cuadrado:** $\sum_{j=0}^{\infty} \theta_j^2 < \infty$
- **Sumabilidad absoluta:** $\sum_{j=0}^{\infty} |\theta_j| < \infty$

Esto, implica que cualquier proceso $MA(q)$ de orden finito, es absolutamente sumable y por tanto, es débilmente estacionario.

1.4 Tipos de procesos: $AR(1)$

Es un tipo de proceso que depende del primer rezago de la serie. Y ese rezago, es lo que nos mide la inercia. Hay algo que ocurrió en el período anterior, pero que nos ayuda a predecir lo que ocurrirá el día de hoy.

$$y_t = \phi y_{t-1} + u_t, \quad u_t \sim WN(0, \sigma^2)$$

Note que por ejemplo, un proceso $MA(q)$ no es ruido blanco, dado que hay correlación distinta de 0 entre los primeros q rezagos. En un $AR(1)$, como el proceso depende de sí mismo rezagado, entonces implícitamente estamos diciendo que depende de todo lo que ha pasado hasta la fecha. Así, tenemos que:

$$y_t = y_{-1} + \sum_{i=0}^t \phi \hat{u}_{t-i} \Rightarrow \frac{\partial y_{t+j}}{\partial u_t} = \phi^j$$

Lo que nos indica que la serie depende de dos cosas: el estado inicial, y de todo lo ocurrido desde ese estado hasta la fecha. Así, vemos que ϕ es el parámetro que determinará el comportamiento de la serie, pudiendo tener los siguientes comportamientos:

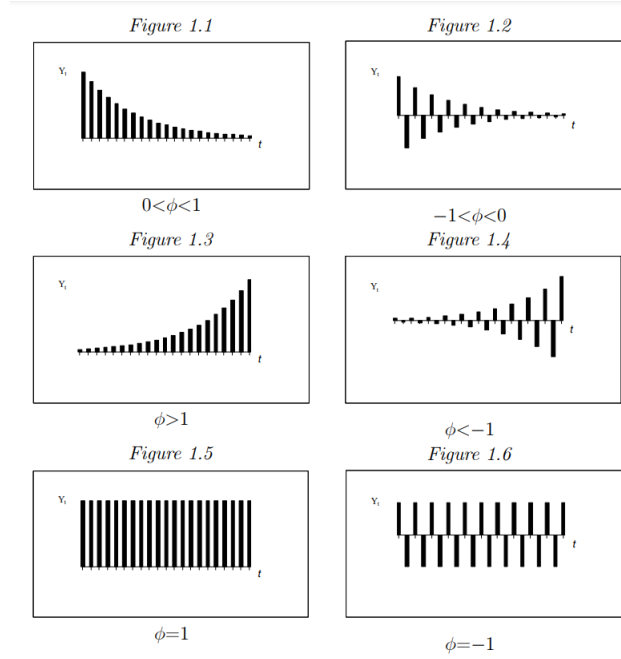


Figure 1: Caption

Donde en las figuras, podemos ver gráficamente las funciones de impulso respuesta a diferentes valores de ϕ . Lo anterior, se puede resumir en lo siguiente:

- $0 < |\phi| < 1$

Cuando $j \rightarrow \infty \Rightarrow |\phi^j| \rightarrow 0$. Luego, los errores se van disipando en el tiempo.

- $|\phi| > 1$

Entonces, el error va influyendo cada vez más en la medida que $j \rightarrow \infty$.

- $|\phi| = 1$

Entonces, tenemos que los shocks son igual de importantes en cada período. Luego, un shock transitorio tiene efectos permanente. Además, esto implica que el proceso tiene una raíz unitaria.

Siguiendo este último proceso, cuando tenemos $|\phi| = 1$, se dice que el proceso es un random walk. Además, en este proceso, el valor esperado en $t+j$, es la misma esperanza en t . Ahora bien, es probable que la variable no tenga el mismo valor en esos períodos, dado que los shocks influyen de manera permanente.

Ahora, veremos cuáles son los momentos del proceso. Además, haremos diferencias entre los momentos condicionales, y los momentos incondicionales de los modelos. Primero, veremos los momentos condicionales, suponiendo la siguiente estructura para el AR(1):

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

Asumiremos que se cumplen las condiciones de estacionariedad débil.

- **Esperanza Condicional:**

$$E(y_t|y_{t-1}) = c + \phi y_{t-1}$$

.

Esto es, la persistencia del rezago nos da información relevante acerca de dónde estaremos en el próximo período.

- **Varianza Condicional:**

$$V(y_t|y_{t-1}) = \sigma^2$$

Lo anterior se da debido a que y_{t-1} es conocido en t , y por tanto no puede variar.

- **Esperanza Incondicional:**

$$E(y_t) \equiv \mu = \frac{c}{1 - \phi}$$

Donde usamos el hecho que $E(y_t) = E(y_{t-1})$ debido al supuesto de DE.

- **Varianza Incondicional:**

$$\gamma_0 = E(y_t - \mu)^2 = \frac{\sigma^2}{1 - \phi^2}$$

Para resolver esto, usamos la propiedad de que la serie auxiliar $x_t = y_t - \mu$ tiene la misma varianza que y_t : $y_t - \mu = \phi(y_{t-1} - \mu) + u_t$. Además, comienzan a aparecer cosas interesantes. Para que la varianza esté bien definida, se debe cumplir que $0 < |\phi| < 1$. Además, si $\phi \rightarrow 1$, entonces la varianza va a ir creciendo indefinidamente.

- **Autocovarianzas:**

$$E(y_t - \mu)(y_{t-1} - \mu) = \phi \gamma_0$$

Luego, el signo de la autocovarianza dependerá del signo de ϕ . Además, esto implica que la autocorrelación será igual a $\rho_1 = \phi$. Por inducción, tendremos que el resto de las autocovarianzas serán:

$$E(y_t - \mu)(y_{t-j} - \mu) = \phi^j \gamma_0, \quad \rho_j = \phi^j$$

Así, note que la autocorrelación de orden j es igual a la función de impulso respuesta de orden j . Esto no necesariamente es cierto siempre, basta ver los procesos MA.

1.5 Tipos de procesos: $AR(p)$

Generalizaremos lo visto en la sección pasada. Ahora, la serie dependerá de un rezago de orden p de sí misma, además del ruido blanco.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

Ahora, mostraremos que todo proceso autorregresivo de orden p puede ser representado como un proceso de primer orden. La idea es definir las siguientes matrices:

$$\underset{(p \times 1)}{\mathcal{E}} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} \quad \underset{(p \times p)}{\mathbf{F}} = \begin{bmatrix} \Phi' & \\ & I \\ & & \mathbf{0} \end{bmatrix}_{\substack{(p-1) \times (p-1) \\ 1 \times (p-1)}}, \quad \underset{(p \times 1)}{\mathbf{v}} = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \underset{(p \times 1)}{\Phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

Y, después el proceso puede ser escrito de la siguiente manera:

$$\mathcal{E}_t = F \mathcal{E}_{t-1} + v_t = F^{t+1} \mathcal{E}_{-1} + \sum_{i=0}^t F^i v_{t-i}$$

Y note que tenemos la misma idea que en un proceso $AR(1)$, donde la serie depende del estado inicial, y de todos los shocks que han habido hasta la fecha. Donde tenemos por ejemplo para un $AR(5)$

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ y_{t-4} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \phi_4 & \phi_5 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ y_{t-4} \\ y_{t-5} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Donde es importante notar que lo que describe y_t en el ejemplo, sería el proceso solo. El resto se utiliza solamente por contabilidad. La representación realizada es un caso particular de algo más general, llamada representación estado espacio. Note que expresada de esta manera, la matriz F va a ser la responsable de caracterizar la dinámica que seguirá esta serie. Así, que debemos estudiarla con atención.

Ahora, veremos como definir si el proceso es débilmente estacionario o no. Para eso, es necesario obtener los eigenvalues de la matriz F , es decir encontrar valores de λ que satisfacen la siguiente condición:

$$\det |F - \lambda I_p| = 0$$

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_{p-1} \lambda - \phi_p = 0$$

Note que los eigenvalues pueden ser Reales o Complejos. Posteriormente, debemos verificar una de las siguientes opciones (equivalentes):

- Eigenvalues están todos dentro del círculo unitario (módulo menor a 1). Se puede verificar solamente evaluando si el máximo de ellos está dentro del círculo unitario.
- Raíces de

$$1 - \phi_1 Z - \cdots - \phi_{p-1} Z^{p-1} - \phi_p Z^p = 0$$

deben ser mayores a 1, verificable solamente considerando que el mínimo sea mayor a 1.

Para evaluar lo anterior, se debe recordar que el módulo de un número complejo $a+bi$ es igual a $|a+bi| = \sqrt{a^2 + b^2}$. Luego, para que esté dentro o fuera del círculo unitario, deben verificar que $|a+bi| \leq 1$. Si hay al menos 1 eigenvalue o raíz iguales a 1, entonces tendríamos una raíz unitaria, concepto que será estudiado más adelante.

Sí el proceso es DE, entonces tendremos los siguientes momentos incondicionales:

$$\mu = \frac{c}{1 - \sum_{i=1}^p \phi_i}$$

$$\gamma_j = \begin{cases} \phi_1 \gamma_{j-1} + \dots + \phi_p \gamma_{j-p} & j = 1, 2, \dots \\ \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma^2 & j = 0 \end{cases}$$

Note que en este caso, la varianza depende de hasta la autocovarianza de orden p . Así, para encontrar los valores de cada una, es necesario resolver un sistema de ecuaciones conocido como Yule-Walker:

$$\rho_j = \phi_1 \rho_{j-1} + \dots + \phi_p \rho_{j-p} \quad j = 1, 2, \dots$$

Este sistema describe como se generan las autocovarianzas y autocorrelaciones de un proceso autoregresivo de orden p .

Ahora, veremos como responde la serie y_t ante shocks transitorios no esperados. Recordando que

$$\varepsilon_t = F^{t+1} \varepsilon_{-1} + \sum_{i=0}^t F^i v_{t-i}$$

Entonces sabemos que la primera fila de esa matriz nos entrega las características de la variable en t . Por tanto, tenemos que:

$$y_t = \sum_{i=1}^p f_{1i}^{(t+1)} y_{-1} + \sum_{i=0}^t f_{11}^{(i)} u_i$$

Es decir la variable en el día de hoy la podemos explicar como una función de la condición inicial y_{-1} y todos los shocks desde esa fecha hasta t . En primera parte, estamos sumando la primera fila de la matriz F^{t+1} , mientras que en la segunda parte nos está diciendo que debemos sumar el primer elemento de la cada matriz F^i ponderado por u_i . Así, es fácil comprobar que las IRF de este proceso tendrán la siguiente forma:

$$\frac{\partial y_{t+j}}{\partial u_t} = f_{11}^{(j)}$$

De todas formas, es posible hacer una descomposición triangular para obtener la IRE, siguiendo el siguiente ejercicio. Si es que los eigenvalues son todos diferentes y t_j es el eigenvector j , entonces podemos descomponer la matriz F de la siguiente manera:

$$\underset{(p \times p)}{F} = \underset{(p \times p)}{T} \underset{(p \times p)}{\Lambda} \underset{(p \times p)}{T^{-1}}$$

Donde tenemos

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}, \quad T = [t_1 \quad t_2 \quad \dots \quad t_p]$$

Donde note que los eigenvalues y los eigenvectors deben estar ordenados de manera correcta. Debido a como está descompuesta la matriz, sabemos que F^j tendrá la siguiente estructura:

$$F^j = T \Lambda^j T^{-1}$$

Evaluando la primera parte de la multiplicación, tenemos que

$$F^j = \begin{bmatrix} t_{11}\lambda_j & \cdots & t_{1p}\lambda_p^j \\ \vdots & \ddots & \vdots \\ t_{p1}\lambda_j & \cdots & t_{pp}\lambda_p^j \end{bmatrix} \begin{bmatrix} t^{11} & \cdots & t^{1p} \\ \vdots & \ddots & \vdots \\ t^{p1} & \cdots & t^{pp} \end{bmatrix}$$

Por lo tanto, tendremos que:

$$f_{11}^{(j)} = [t_{11}t^{11}] \lambda_1^j + \cdots + [t_{1p}t^{p1}] \lambda_p^j = \sum_{i=1}^p c_i \lambda_i^j$$

Donde $c_i = [t_{1i}t^{i1}]$, $\sum_{i=1}^p c_i = 1$ dado que los componenets son de matrices inversas, y por tanto, el primer elemento de la multiplicación entre ellas debe ser igual a 1. Por último, es posible demostrar que cada uno de los componentes c_i , puede representarse como una función solamente de los eigenvalues de la matriz F :

$$\text{Si } \lambda_i \neq \lambda_j, \quad \forall i \neq j \Rightarrow c_i = \frac{\lambda_i^{p-1}}{\prod_{k \neq i} (\lambda_i - \lambda_k)}$$

Además, si tenemos que $\lambda_i \in \mathbb{R}, \forall i$, sea λ_1 el eigenvalue con mayor valor absoluto. Entonces

- $|\lambda_1| < 1$ la respuesta es estable, y el proceso es DE
- $|\lambda_1| = 1$ el proceso es no estacionario, y tiene raíz unitaria.
- $|\lambda_1| > 1$ la respuesta es explosiva

1.6 Invertibilidad

Diremos que un proceso MA es invertible si es que puede ser escrito como una $AR(\infty)$ Por ejemplo, para un $MA(1)$ con $|\theta| < 1$

$$y_t = (1 + \theta L)u_t \iff \frac{y_t}{1 + \theta L} = u_t \iff \sum_{i=0}^{\infty} (-\theta)^i y_{t-i} = u_t$$

Donde estamos usando polinomios de rezago para realizar estas operaciones. Si $|\theta| > 1$, en teoría no podríamos hacer la conversión, en tanto la sumatoria no convergería. Sin embargo, es posible realizar un pequeño truco. Definiendo $\delta = 1/\theta$ y $\sigma_v^2 = \sigma^2 \theta^2$, entonces tendríamos un proceso que sí es invertible, y que tiene las mismas autocorrelaciones y autocovarianzas. El único caso en el que no podremos hacer esta transformación, es en el caso en el que haya una raíz unitaria.

Si tuvieramos un $MA(q)$, se debe construir una matriz F del proceso de media móvil, Si todos los eigenvalues de F son menores a 1, el proceso es invertible. Si algún eigenvalue es mayor a 1, se podrá reparametrizar la serie a una observacionalmente equivalente, pero que sea invertible. Todo excepto en el caso en que haya una raíz unitaria.

Para el caso de un $AR(1)$ con $|\phi| < 1$, tenemos que

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

$$y_t = \frac{u_t}{1 - \phi L}$$

$$y_t = \sum_{i=0}^{\infty} \phi^i u_{t-i}$$

Luego, un $AR(1)$ DE puede ser descrito como un $MA(\infty)$

1.7 Tipos de procesos: $ARMA(p, q)$

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q}$$

Luego, podemos expresar este proceso en términos de polinomios de rezagos, y en caso de que todos los eigenvalues del proceso de la parte AR estén dentro del círculo unitario, tendremos:

$$y_t = \mu + \frac{1 + \theta_1 L + \dots + \theta_q L^q}{1 - \phi_1 L - \dots - \phi_p L^p}, \mu = \frac{c}{1 - \phi_1 - \dots - \phi_p}$$

Para el resto de los momentos, habría que tener ciertas consideraciones.

- $\gamma_j, j > q$ son idénticas a un proceso $AR(p)$
- $\gamma_j, j \leq q$ son más complejas debido a la parte MA.

Además, note que con los polinomios de rezago, es posible simplificar el modelo lo más posible. Así, hay que tener cuidado con la sobre parametrización:

$$y_t = u_t \iff (1 - \delta L)y_t = (1 - \delta L)u_t$$

Esto es, si es que hay alguna raíz compartida entre el componente de MA y AR, entonces podemos bajar un orden en ambas partes del proceso: $ARMA(p, q) = ARMA(p-1, q-1)$

1.8 Estimación AR(1)

Ahora, explicaremos cuáles son las técnicas generalmente usadas para la estimación de modelos ARMA. Como nota, no siempre sabemos cuál es la estructura real de los datos, por lo que ni siquiera sabemos el orden de ARMA que debemos estimar. De todas formas, asumiendo que sí conocemos la estructura, presentaremos formas de estimación. Además, usaremos como supuesto que u_t distribuye normal. La gracia de este supuesto, es que nos permite saber cuál es el proceso que genera los datos de u_t , con lo que podremos utilizar máxima verosimilitud. Si no supieramos cuál es la distribución de u_t , se puede usar cuasi-máximaverosimilitud. Comenzaremos con un $AR(1)$.

$$y_t = c + \phi y_{t-1} + u_t, u_t \sim \mathcal{N}(0, \sigma^2)$$

Las formas de estimar esto por ML, son dos. El método exacto, y el método condicional. La diferencia entre estas dos estimaciones, es cómo lidiar con la observación que se podría perder. Esto dado que para estimar, necesitamos tanto y_t como y_{t-1} . Luego, como necesitamos 1 rezago, se pierde 1 dato a la hora de estimar. Como repaso, tenemos que la función de verosimilitud nos plantea una pregunta inversa a la que responde la función de distribución. Dados los datos, cuáles son los parámetros que con mayor probabilidad, podrían haber generado la serie observada?. Por último, es necesario que las funciones de MV, deben estar expresadas solamente en función de observables y parámetros.

- **Método Exacto:** Para el método exacto, asumimos que antes de la primera observación, la serie se encontraba en sus momentos incondicionales. Por lo tanto, **es necesario que la serie sea estacionaria** (al menos débilmente) para poder aplicar este método de estimación.

De esta forma, para las observaciones tales que $t \geq 2$, tenemos que:

$$l(c, \phi, \sigma^2; y_t | y_{t-1}) \equiv f(y_t | y_{t-1}; c, \phi, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}$$

Donde note que usamos los momentos condicionales de la serie a la hora de despejar el u_t . Esto, dado que y_{t-1} es conocido. Mientras que para la primera observación, sería:

$$l(c, \phi, \sigma^2; y_1) \equiv f(y_1; c, \phi, \sigma) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{1-\phi^2}}} \exp - \frac{(y_1 - \frac{c}{1-\phi})^2}{2 \frac{\sigma^2}{1-\phi^2}}$$

Luego, la Log-Likelihood del problema quedaría de la siguiente manera:

$$\mathcal{L}(c, \phi, \sigma^2; y_T, \dots, y_1) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\frac{\sigma^2}{1-\phi^2} \right) - \frac{\left(y_1 - \frac{c}{1-\phi} \right)^2}{2 \frac{\sigma^2}{1-\phi^2}} - \frac{T-1}{2} \ln(\sigma^2) - \sum_{t=2}^T \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}$$

Finalmente, para poder obtener los estimadores, se debe derivar la expresión con respecto a los parámetros, e igualar a 0. Como lo anterior da lugar a un sistema de ecuaciones no lineales, generalmente no es posible encontrar una solución analítica al problema. De esta forma, para poder estimar esta función, se deberá acudir a métodos numéricos (Newton-Raphson u otros).

- **Método Condicional:** Se llama de esta forma, dado que utilizamos solamente los datos en los que sabemos que podemos condicionar en la observación de los rezagos.

Así, tendremos que la función log-likelihood en este caso tendrá la siguiente forma:

$$\mathcal{L}(c, \phi, \sigma^2; y_T, \dots, y_1) = \sum_{t \geq 2} \ln l(c, \phi, \sigma^2; y_t | y_{t-1}) = -\frac{T-1}{2} \ln(\sigma^2) - \sum_{t=2}^T \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}$$

Y, dado lo anterior, sabemos que maximizar \mathcal{L} es lo mismo que minimizar la suma de los cuadrados de los errores $\sum (y_t - c - \phi y_{t-1})^2$. Por tanto, el estimador MLE es el mismo que el de OLS.

$$Y)X\beta + u$$

$$\begin{matrix} Y \\ (T-1) \times 1 \end{matrix} = \begin{bmatrix} y_2 \\ \vdots \\ y_T \end{bmatrix}, \begin{matrix} X \\ (T-1) \times 2 \end{matrix} = \begin{bmatrix} 1 & y_1 \\ \vdots & \vdots \\ 1 & y_{T-1} \end{bmatrix}, \begin{matrix} u \\ (T-1) \times 1 \end{matrix} = \begin{bmatrix} u_2 \\ \vdots \\ u_T \end{bmatrix}, \begin{matrix} \beta \\ 2 \times 1 \end{matrix} = \begin{bmatrix} c \\ \phi \end{bmatrix}$$

Luego, el estimador OLS será

$$\hat{\beta} = (X'X)^{-1} (X'Y), \quad \hat{\sigma}^2 = (T-1)^{-1} \hat{u}'\hat{u} = \sum_{t=2}^T \frac{(y_t - \hat{c} - \hat{\phi} y_{t-1})^2}{T-1}$$

Para este caso, no es necesario que la serie sea estacionaria. Por último, está demostrado que la estimación condicional y la exacta, para cuando el proceso es débilmente estacionario, son asintóticamente equivalentes. De esta forma, la única situación en la que podía valer la pena estimar mediante la exacta, es cuando tenemos un tamaño de muestra relativamente pequeño, y por tanto, se necesita utilizar todos los datos disponibles. Además, como en los métodos numéricos es necesaria una condición inicial para estimar los parámetros, es posible utilizar los estimadores de OLS, aumentando la velocidad de estimación.

1.9 Estimación AR(p)

En esencia, la estimación AR(p) es igual a la del AR(1). Para ver el término condicional, como ahora hay más rezagos, habría una pérdida de más observaciones para poder realizar la estimación. Así:

$$\mathcal{L}(c, \Phi, \sigma^2; y_T, \dots, y_1) = \sum_{t=2}^T \ln l(c, \Phi, \sigma^2; y_t | y_{t-1}) = -\frac{T-1}{2} \ln(\sigma^2) - \sum_{t=2}^T \frac{(y_t - c - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p})^2}{2\sigma^2}$$

Y nuevamente, maximizar \mathcal{L} es lo mismo que minimizar la suma del cuadrado de los errores $\sum_{t=2}^T (y_t - c - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p})^2$. Por tanto, el estimador MLE es el mismo que el OLS:

$$\hat{\beta} = (X'X)^{-1} (X'Y), \quad \hat{\sigma}^2 = (T-p)^{-1} \hat{u}'\hat{u} = \sum_{t=p}^T \frac{(y_t - \hat{c} - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p})^2}{T-p}$$

Para el caso exacto (incondicional), es básicamente el mismo ejercicio de generalización. Así, para las p primeras observaciones, se deben poner los momentos condicionales. Note eso sí, que en la medida que va pasando de rezago en rezago, ya tiene una realización de la serie. Así, se van necesitando cada vez menos términos iguales a la esperanza incondicional. Lo anterior, implica que el sistema de ecuaciones a resolver para ese caso es muy no lineal. Casi no hay ventajas de estimar de esa manera, salvo algunos casos puntuales.

1.10 Estimación MA(1)

$$y_t = \mu + u_t + \theta u_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2)$$

Esta estimación es más compleja, dada la naturaleza de la serie. Para poder escribir la función de máxima verosimilitud, es necesario escribir todo en función de observables y parámetros. Pero u_t no es observable. Para esto, la técnica para estimar es la siguiente

- **Estimación condicional:** Comenzaremos la estimación condicional en que $u_1 = 0$. De esta forma, podremos ir recuperando recursivamente el resto de los u_t , dejándolos en función de variables observables.

Si $u_1 = 0$, entonces:

$$\begin{aligned} (y_1 | u_0 = 0) &\sim \mathcal{N}(\mu, \sigma^2) \Rightarrow l(\mu, \theta, \sigma^2; y_1 | u_0 = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y_1 - \mu)^2}{2\sigma^2} \\ u_1 &= y_1 - \mu \Rightarrow l(\mu, \theta, \sigma^2; y_2 | y_1, u_0 = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y_2 - \mu - \theta u_1)^2}{2\sigma^2} \\ u_2 &= y_2 - \mu - \theta u_1 \Rightarrow l(\mu, \theta, \sigma^2; y_3 | y_2, u_0 = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y_3 - \mu - \theta u_2)^2}{2\sigma^2} \\ u_t &= y_t - \mu - \theta u_{t-1} \Rightarrow l(\mu, \theta, \sigma^2; y_t | y_{t-1}, \dots, y_1, u_0 = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{u_t^2}{2\sigma^2} \end{aligned}$$

Así, la función de loglikelihood quedaría de la siguiente manera

$$\mathcal{L} = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \sum_{t=1}^T \frac{u_t^2}{2\sigma^2}$$

De esta forma, las ecuaciones para resolver la log-likelihood son muy no lineales, dado que recursivamente se encuentran los parámetros en cada iteración. Como necesitamos una condición inicial, en caso de que el proceso sea invertible, entonces la condición inicial irá dejando de ser relevante. En caso de que se obtenga un valor de $|\theta| > 1$, entonces usamos el proceso observacionalmente equivalente, discutido arriba.,

1.11 Estimación MA(q)

Sigue el mismo espíritu que para estimar un MA(1). Así, tenemos que:

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q}, \quad u_t \sim \mathcal{N}(0, \sigma^2)$$

Solamente que se asumirá que $u_0 = u_{-1} = u_{1-q} = 0$

De esta misma manera, seguiremos el mismo proceso anterior

$$u_t = y_t - \mu - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$$

Lo que nos lleva a una función log-likelihood muy similar a la anterior

$$\mathcal{L} = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \sum_{t=1}^T \frac{u_t^2}{2\sigma^2}$$

Algunas de las limitaciones son que

- Nuevamente, es muy no lineal en los parámetros
- Requiere que las raíces del MA estén fuera del círculo unitario (que sea invertible). Esto es necesario para que las condiciones iniciales sean irrelevantes si el tamaño de la muestra es grande.
- Si no, se puede utilizar un proceso observacionalmente equivalente.

Para el caso de la estimación incondicional, es aún más compleja, pero no requiere que las raíces estén fuera del círculo unitario.

1.12 Estimación ARMA(p,q)

Para estimar ese proceso, debemos combinar la parte de media móvil con la forma de estimación de un AR(p).

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q}$$

Así, para estimar la condicional, necesitamos que $u_0 = u_{-1} = u_{1-q} = 0$, pero además, debemos tomar en cuenta la parte autorregresiva, por lo que solo podemos partir desde la observación p en adelante. Así, sabemos que cada error se construirá de la siguiente manera:

$$u_t = y_t - c - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}, \quad t \geq p+1$$

Lo anterior, da lugar a la siguiente log-likelihood

$$\mathcal{L} = -\frac{T-p}{2} \ln(2\pi) - \frac{T-p}{2} \ln(\sigma^2) - \sum_{t=p+1}^T \frac{u_t^2}{2\sigma^2}$$

Algunas consideraciones para la estimación serían:

- Nuevamente, las raíces de la parte MA deben estar fuera del círculo unitario, para que las condiciones iniciales se vayan disipando en el tiempo.

- Para estimar, debemos utilizar métodos numéricos.

ARMAX: ARMA donde también podemos poner variables exógenas que no están modeladas.

1.13 Propiedades asintóticas

En esta sección, asumiendo que sabemos cuál es el verdadero DGP de las series, y que hemos estimado sus parámetros, nos gustaría saber cuáles son las propiedades de los estimadores. Hay varias maneras de evaluar estas propiedades. Algunas, se apoyarán en la teoría asintótica, intentando responder dos preguntas:

- ¿Cuán alejado está el estimador del verdadero vector de parámetros?

Para responder a esta pregunta, se utilizarán variantes de leyes de grandes números. Diremos que un estimador es débilmente consistente, si converge en probabilidad al verdadero vector de parámetros $\hat{\beta} \xrightarrow{P} \beta$, y diremos que es fuertemente consistente si es que converge de manera casi segura al verdadero vector de parámetros $\hat{\beta} \xrightarrow{CS} \beta$

- ¿Cómo podemos hacer inferencia respecto del verdadero valor de los parámetros, si queremos testear alguna hipótesis en particular?

Para este caso, nos apoyaremos en variantes del Teorema Central del Límite para poder hacer inferencia. Habrían otras formas para responder a estas preguntas, como experimentos de MonteCarlo o refinamientos asintóticos.

1.14 Propiedades asintóticas: AR(p)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

Primero, veremos las propiedades asintóticas de un proceso de orden p en general. Además, asumiremos que el proceso es DE. Más adelante, se generalizará aún más. Mediante la estimación de Máxima Verosimilitud condicional, sabemos que los estimadores son los mismos que los que se obtienen mediante un MCO:

$$Y = X\beta + u$$

Donde Y es nuestra variable en t , y X sería una matriz de rezagos de la variable dependiente. Para eso, analizaremos el siguiente teorema que nos plantea las propiedades asintóticas de este estimador.

Teorema 1. Si $u_t \sim iid(0, \sigma^2)$ y son independientes de cualquier rezago de y_t (note que no se necesita normalidad de los errores), y además, requerimos que el proceso sea absolutamente sumable (DE). Entonces, tenemos que

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2 \Sigma_p^{-1})$$

, donde además tenemos que

$$\hat{\sigma}^2 = (T - p)^{-1} (Y - X\hat{\beta})' (Y - X\hat{\beta}) \xrightarrow{P} \sigma^2$$

$$(T - p)^{-1} X'X \xrightarrow{P} \sigma^2$$

Donde $\hat{\sigma}^2$ es el estimador de la varianza de u_t , y $(T - p)^{-1} X'X$ es el estimador de la matriz de covarianzas del proceso.

Luego, tenemos que la teoría asintótica de OLS prevalece para el caso donde el proceso es estacionario.

1.15 Demostración propiedades asintóticas de un AR(1)

Supongamos que tenemos el siguiente proceso:

$$y_t = \phi y_{t-1} + u_t, u_t \sim iid(0, \sigma^2), |\phi| < 1$$

Ahora, demostraremos que el estimador es consistente y asintóticamente normal.

- **Consistencia.**

Note que en primer lugar, asumiendo SPG que $c = 0$

$$y_t = \begin{bmatrix} y_2 \\ \vdots \\ y_T \end{bmatrix}, X = \begin{bmatrix} y_1 \\ \vdots \\ y_{T-1} \end{bmatrix}$$

Luego, el estimador de OLS en este caso será

$$\hat{\beta} = \hat{\phi} = (X'X)^{-1} X'Y = \frac{\sum_{t=2}^T y_{t-1} y_t}{\sum_{t=2}^T y_t^2}$$

Sustituyendo $y_t = \phi y_{t-1} + u_t$, y con un poco de álgebra, tenemos que:

$$\hat{\phi} - \phi = \frac{\sum_{t=2}^T y_{t-1} u_t}{\sum_{t=2}^T y_t^2} = \frac{\sum_{t=2}^T y_{t-1} u_t / T}{\sum_{t=2}^T y_t^2 / T}$$

Para evaluar la consistencia del estimador, veremos a qué converge el numerador y el denominador de la expresión de arriba. Así, sabremos que la convergencia de la división será la división de los límites si el denominador es $\neq 0$. Ahora, utilizaremos leyes de grandes números, para ver si los promedios muestrales logran converger a esperanzas poblacionales. Por eso, dividimos por T tanto el numerador como el denominador. A continuación, mostraremos que el numerador converge en probabilidad a 0, el denominador a un valor estrictamente positivo, por lo que la fracción converge en probabilidad a 0.

Para el numerador, demostraremos utilizadno Tchevishev, y utilizando además, que convergencia en media cuadrada implica convergencia en probabilidad. Así, mostraremos que $NUM = \frac{\sum_{t=2}^T y_{t-1} u_t / T}{\sum_{t=2}^T y_t^2 / T} \xrightarrow{SM} 0 \Rightarrow NUM \xrightarrow{P} 0$

Esto es, hay que demostrar que

$$E \left[\frac{\sum_{t=2}^T y_{t-1} u_t / T}{\sum_{t=2}^T y_t^2 / T} \right]^2 \rightarrow 0, \text{ if } T \rightarrow \infty$$

Así:

$$E \left[\frac{\sum_{t=2}^T y_{t-1} u_t / T}{\sum_{t=2}^T y_t^2 / T} \right]^2 = \frac{1}{T^2} E(\sum y_{t-1} u_t)^2 = \frac{1}{T^2} E(\sum v_t)^2$$

Donde $y_{t-1} u_t = v_t$. Note que como u_t y y_{t-1} son independientes, entonces $E(v_t) = E(u_t)E(y_{t-1}) = 0$, y además $V(v_t) = E(u_t^2 y_{t-1}^2) = E(u_t^2)E(y_{t-1}^2)$. Como conocemos cuáles son las varianzas incondicionales de ambos procesos, tenemos que finalmente $V(v_t) = \frac{\sigma^4}{1-\phi^2}$. Aplicando lo anterior a la suma del cuadrado de la media, tenemos:

$$E \left[\frac{\sum_{t=2}^T y_{t-1} u_t / T}{\sum_{t=2}^T y_t^2 / T} \right]^2 = \frac{1}{T^2} E(\sum v_t)^2 = \frac{1}{T^2} \left[E \left(\sum y_{t-1}^2 u_t^2 \right) + 2E \left(\sum_t \sum_j y_{t-1} u_t y_{t-j} u_{t-j+1} \right) \right]$$

Lo importante de este caso, es notar que como los errores son ruido blanco, no están correlacionados ni con errores de otros períodos, ni con rezagos de la serie. Luego, la esperanza de esa doble sumatoria es igual a 0.

$$E \left[\frac{\sum_{t=2}^T y_{t-1} u_t / T}{\sum_{t=2}^T y_t^2 / T} \right]^2 = \frac{1}{T^2} E \left(\sum v_t \right)^2 = \frac{1}{T^2} \sum V(y_{t-1} u_t) = \frac{T-1}{T^2} \frac{\sigma^4}{1-\phi^2}$$

Luego, tomando el límite:

$$\lim_{T \rightarrow \infty} NUM = \lim_{T \rightarrow \infty} \frac{T-1}{T^2} \frac{\sigma^4}{1-\phi^2} = 0$$

Con esto, tenemos que $NUM \xrightarrow{SM} 0 \Rightarrow NUM \xrightarrow{P} 0$

Ahora, vamos con el denominador, el que tenía la siguiente forma:

$$DEN = \frac{\sum y_{t-1}^2}{T}$$

Para lo anterior, usaremos dos resultados previos: $\sum y_{t-1} u_t = \sum y_{t-1} (y_t - \phi y_{t-1}) \xrightarrow{P} 0$ llamada condición *a*), y por otro lado, $\frac{1}{T} \sum u_t^2 = \frac{1}{T} \sum (y_t - \phi y_{t-1})^2 \xrightarrow{AS} \sigma^2$ llamada condición *b*), por leyes de grandes números de Kolmogorov.

A continuación, tomemos $b) + 2\phi a)$

$$\begin{aligned} \frac{1}{T} \left[\sum (y_t - \phi y_{t-1})^2 + 2\phi \sum y_{t-1} (y_t - \phi y_{t-1}) \right] &= \frac{1}{T} \left[\sum y_t^2 - 2\phi \sum y_t y_{t-1} + \phi^2 \sum y_{t-1}^2 + 2\phi \sum y_t y_{t-1} - 2\phi^2 \sum y_{t-1}^2 \right] \\ &= \frac{1}{T} \left[\sum y_t^2 - \phi^2 \sum y_{t-1}^2 \right] \xrightarrow{P} \sigma^2 \end{aligned}$$

Así, factorizando de manera conveniente, tenemos que:

$$= \frac{1}{T} \left[(1 - \phi^2) \sum_{t=2}^T y_t^2 + y_T^2 - \phi^2 y_1^2 \right] \xrightarrow{P} \sigma^2$$

Como el último termino y el primer término son finitos, cuando interactúan con T^{-1} se van a 0. Así, tenemos que

$$\frac{1}{T} (1 - \phi^2) \sum y_t^2 \xrightarrow{P} \sigma^2 \iff \frac{1}{T} \sum y_t^2 \xrightarrow{P} \frac{\sigma^2}{1 - \phi^2}$$

Por tanto, tenemos que

$$DEN \xrightarrow{P} \gamma_0 = \frac{\sigma^2}{1 - \phi^2}$$

Luego, esto implica que

$$\hat{\phi} - \phi = \frac{NUM}{DEN} \xrightarrow{P} 0$$

Donde utilizamos Slutsky en la última parte. Como corolario, tenemos que $\hat{\phi} \xrightarrow{P} \phi$

- **Normalidad Asintótica.**

Para poder saber cuál es la distribución asintótica, debemos hacer un cambio en la escala para saber la distribución de los parámetros. Así, tomamos la siguiente expresión:

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

Nuevamente separando numerador de denominador, tenemos que:

$$NUM = T^{-\frac{1}{2}} \sum y_{t-1} u_t, \quad DEN = T^{-1} \sum y_{t-1}^2 \xrightarrow{P} \frac{\sigma^2}{1 - \phi^2}$$

Como ya sabemos a qué converge el denominador, nos centraremos en el numerador. Nuevamente, tomaremos la variable auxiliar v_t , donde $E(v_t) = 0$, $V(v_t) = \frac{\sigma^4}{1 - \phi^2}$. Ahora, en este caso debemos aplicar el Teorema Central del Límite 1. Con eso, podemos obtener que

$$NUM = T^{-\frac{1}{2}} \sum y_{t-1} u_t \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^4}{1 - \phi^2}\right)$$

Finalmente uniendo ambos términos tenemos que

$$\sqrt{T}(\hat{\phi} - \phi) = \frac{NUM}{DEN} \xrightarrow{D} \mathcal{N}(0, 1 - \phi^2)$$

1.16 Distribución asintótica ARMA(p,q)

La distribución asintótica de un proceso $ARMA(p, q)$ es muy similar a la que vimos antes. Así, tendremos que

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}\left(0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1}\right)$$

Donde

$$A(\theta_0) = -\lim_{T \rightarrow \infty} ET^{-1} \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'} \Big|_{\theta_0}$$

$$B(\theta_0) = \lim_{T \rightarrow \infty} ET^{-1} \frac{\partial \mathcal{L}}{\partial \theta} \Big|_{\theta_0} \frac{\partial \mathcal{L}}{\partial \theta'} \Big|_{\theta_0}$$

Donde si el modelo está bien especificado (esto es, conocemos el DGP y a su vez, los errores distribuyen normal), tenemos que $A(\theta_0) = B(\theta_0)$, por lo que

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}\left(0, A(\theta_0)^{-1}\right)$$

Y finalmente, tenemos que

$$E(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0) = \left[-\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'} \Big|_{\hat{\theta}} \right]^{-1}$$

y donde los test LRT y LM se puede llevar de manera usual.

2 Econometría Inglesa

2.1 Introducción

Econometría requiere interacción de dos cosas, teoría económica y datos estadísticos. Cuando se hace un modelamiento, se intenta hacer un match entre teoría y los datos en un marco formal cuantitativo.

Además, notar que no todos los modelos nacen iguales. Las razones por las que estos son producidos, son variadas, por lo que se busca usar la econometría para falsear teorías. Otras veces, se busca hacer proyecciones para el futuro, por lo que se busca tener un modelo econométrico que sea confiable a la hora de hacer proyecciones. Se puede usar la econometría para analizar el efecto de políticas. Después de la crítica de Lucas, sabemos que hay ciertas restricciones a tomar en cuenta. Básicamente, se debe tener en claro cuál es la pregunta que se quiere investigar, y a raíz de eso, ver cuál es el mejor enfoque se puede tener en un modelo dada esa pregunta.

Generalmente hay dos enfoques extremos para responder preguntas:

- **Enfoque en la teoría económica:** Hacemos un modelo teórico, y derivamos algunas regularidades que deberían ocurrir, se calibra el modelo y se analizan escenarios contrafactuales. No hay mucha econometría, por lo que es difícil de falsear. Theory dependence: si se cae la teoría, se caen todas las derivaciones provenientes de la misma. Igual de todas formas, recordar que de por sí todos los modelos son falsos por definición, y solo se usan para entregar respuestas satisfactorias respecto de algún tema en particular.
- **Enfoque en los datos:** No se respeta nada la teoría, o se ignora completamente. Por ejemplo, cuando se busca caracterizar regularidades empíricas de los datos, pero no se busca dar una explicación económica profunda acerca de por qué ocurren las cosas. Sample Dependence: Observaciones guían las conclusiones. Cambiando las observaciones, cambian las observaciones. Luego, se puede tener modelos solamente para el status quo, siendo muy sensibles a cambios estructurales.

Econometría inglesa intenta juntar satisfactoriamente estas dos cosas, y además provee maneras rigurosas de testear estos modelos.

2.2 Outline

Estamos interesados en encontrar el DGP de las series. Existe una estructura en la economía, pero no sabemos cuál es. Buscamos buscar una forma satisfactoria de encontrar el proceso que genera los datos. El enfoque propuesto para este fin, se conoce como GETS: General to Specific.

Otro aporte importante es la discusión de endogeneidad. Después de la Crítica de Lucas, se definieron nuevos requisitos que se deben cumplir para que las variables sean exógenas, y habrán diferentes tipos de endogeneidad.

Veremos congruencia de modelos: un listado de requisitos que un modelo debe satisfacer cuando intentamos caracterizar el proceso que sigue una serie. Esto es de lo más importante.

Finalmente, veremos algo de relaciones espurias, y cointegración.

2.3 Enfoque GETS

Siempre deberíamos intentar comenzar de un modelo general para llegar al particular. De otra forma, es posible que estaríamos sesgando nuestro modelo a uno incorrecto, y que no hubieramos llegado en caso de haber comenzando de otra forma. Comenzar de una especificación más general, a algo simple. Así, Henry nos muestra una tipología de modelos que pueden salir a partir del modelo 1 de la figura de abajo (se podría tener algo más general, pero es para hacer el punto).

Typology			
Name	Specification		Restrictions
General	M_1	$y_t = \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t$	None
Static	M_2	$y_t = \beta_1 x_t + u_t$	$\beta_2 = \beta_3 = 0$
Time series	M_3	$y_t = \beta_2 y_{t-1} + \omega_t$	$\beta_1 = \beta_3 = 0$
Differences	M_4	$\Delta y_t = \beta_1 \Delta x_t + \zeta_t$	$\beta_2 = 1, \beta_3 = -\beta_3$
Leading indicator	M_5	$y_t = \beta_3 x_{t-1} + v_t$	$\beta_1 = \beta_2 = 0$
Partial adjustment	M_6	$y_t = \beta_1 x_t + \beta_2 y_{t-1} + \eta_t$	$\beta_3 = 0$
Common factors	M_7	$y_t = \beta_1 x_t + u_t, u_t = \beta_2 u_{t-1} + e_t$	$\beta_3 = -\beta_1 \beta_2$
Distributed lags	M_8	$y_t = \beta_1 x_t + \beta_3 x_{t-1} + \xi_t$	$\beta_2 = 0$
Dead start	M_9	$y_t = \beta_2 y_{t-1} + \beta_3 x_{t-1} + v_t$	$\beta_1 = 0$
EC (homogenous)	M_{10}	$\Delta y_t = \beta_1 \Delta x_t + (\beta_2 - 1)(y - x)_{t-1} + v_t$	$\sum_{i=1}^3 \beta_i = 1$
EC (general)	M_{11}	$\Delta y_t = \beta_1 \Delta x_t + (\beta_2 - 1)(y - \kappa x)_{t-1} + \epsilon_t$	None

Figure 2: Caption

Partiendo del modelo 1, se puede llegar a cualquiera de los modelos siguientes, testeando las restricciones que se encuentran ahí. La moraleja es que se debe tener una mente abierta cuando se comienza por evaluar los modelos. Se debe ser riguroso en como se testean las restricciones para llegar a otro tipo de modelos.

Idea general, es que estamos intentando recuperar cuál es el DGP. Luego, se utiliza la teoría económica para comenzar con un modelo lo más general posible, y luego de eso tratar de obtener una representación más fina de del DGP, la cual debe tener una relación paramétrica simple, que sea estable en el tiempo, que tenga interpretación económica y a su vez, que pueda dar cuenta de resultados de modelos rivales (que pueda encompar).

Las etapas que se nos proponen para lograr el objetivo son las siguientes:

1. **Marginalización:** Utilizando la teoría económica, definir a priori que se considerará y que no.
2. **Condicionamiento:** Se comienza a distinguir respecto de los conjuntos de variables. En principio, todo depende de todo (partir de lo general). Sin embargo, vamos a necesitar distinguir cuáles son las variables que serán endógenas y cuáles serán exógenas, lo cuál será testeable. Teoría económica puede ayudar con la intuición, sin embargo la última palabra la tienen los test.
3. **Forma General Sistemática:** Luego de haber hecho lo anterior, los resultados deberían seguir siendo generales, pero limpios de información que podría no ser relevante.
4. **Simplificar:** Nos gustaría siempre tener un modelo más chico a un modelo más grande (cuando los datos lo permitan). Testear cada reducción
5. **Reparametrizar:** Si dos variables relacionadas tienen un parámetro similar, se puede utilizar el mismo parámetro para ambas, lo que aumenta la eficiencia de la estimación. Testear cada reparametrización.
6. **Reducir**
7. **Comparar con modelos rivales**

2.4 Exogeneidad

No todos los modelos nacen iguales. No a todos los modelos debemos pedir los mismos requerimientos. Criterios de exogeneidad seguirán en general 3 objetivos posibles:

- **Hacer inferencia válida:** Para testear una teoría. Supongamos el siguiente modelo:

$$y_t = \beta x_t + e_t$$

Y por lo tanto, como y_t depende de x_t , y_t es la variable endógena y x_t es la variable exógena. Cuando estamos condicionando en el **valor contemporáneo** de una variable, como el caso del ejemplo, estamos asumiendo que la variable sobre la que condicionamos es exógena (débil) con respecto al vector de parámetros de interés. Para hacer inferencia válida, se necesita que la variable x_t se **exógena débil**.

- **Hacer predicciones:** Una cosa que debería cumplirse en este caso, es que además de débilmente exógena, cumpla con ser **fuertemente exógena**. Esto es una característica específica que se pedirá al modelo. Siguiendo el ejemplo anterior, si se busca hacer proyecciones de y_t en el tiempo, se necesitará saber el valor predicho de x_t en el futuro, por lo que las proyecciones de x_t deben cumplir con ser fuertemente exógenas respecto de y_t .
- **Evaluaciones de Política Contrafactual:** Para este caso, se necesita tener **super exogeneidad**, la cuál será una manera de testear la crítica de Lucas (si cambia la ley de movimientos del estado, deberían cambiar las policy functions o la manera en la que y_t reacciona a x_t). Por ejemplo, si cambia la política fiscal, se podría esperar que cambie la manera en la que los agentes responden ante los impuestos.

Supongamos que tenemos un conjunto de observaciones $\{w_1, w_2, \dots, w_T\}$, con T igual al número de observaciones que tenemos. Nosotros hemos típicamente particionado w_t entre una variable y_t y k variables x_t : $w_t = (y_t, x_t)$ $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^k$. Así, y_t sería nuestra variable endógena y x_t nuestra variable exógena, por lo que necesitamos un criterio para definir este último concepto para verificar si es que se hizo la partición de manera correcta.

Por el teorema de Bayes, tenemos que la distribución conjunta de y_t y x_t se puede escribir como el producto de la condicional de y_t en x_t por la marginal de x_t :

$$f(y_t, x_t | Y_{t-1}, X_{t-1}, \theta) = f(y_t | x_t, Y_{t-1}, X_{t-1}, \theta_1) \cdot f(x_t | Y_{t-1}, X_{t-1}, \theta_2)$$

Donde W_{t-1} son todos los rezagos hasta $t-1$ de la variable w_t . La distribución conjunta nos muestra la probabilidad de observar al mismo tiempo y_t y x_t . En la condicional, se tiene que y_t depende de x_t y no se ven al mismo tiempo. Se tiene una postura, donde se cree que x_t causa a y_t , dado que por teorema de Bayes se podría haber escrito al revés. Lo ideal en este caso es que podamos hacer inferencia sobre los parámetros de $y_t | x_t$ ignorando los de la distribución marginal de x_t . No nos interesa toda la distribución conjunta, sino que una forma para lograr estimar los parámetros de interés dada una función $\psi(\theta)$. Así, lo que nos gustaría tener sería un modelo satisfactorio, reconocer los parámetros de interés y lograr hacer inferencia válida.

Ahora, veremos las definiciones de las categorías de exogeneidad que vimos anteriormente. Antes de comenzar, notar que una variable será exógena siempre respecto de algo (exogeneidad es una categoría relativa). Siempre será exógena respecto del vector de parámetros de interés.

- **Exogeneidad débil:** Se deben cumplir dos condiciones:
 1. Tiene que haber un cohorte secuencial, es decir, que los parámetros de la condicional $y_t | x_t$ (θ_1) y los parámetros de la marginal de x_t (θ_2) tienen que ser libres de variación: No hay manera que haya una asociación entre θ_1 y θ_2 . Si no son libres de variación y θ_2 cambia, entonces también cambiará θ_1 modificando la relación con los parámetros de interés.

2. El vector de parámetros de interés $\psi(\theta)$ debe ser solamente función de θ_1 : $\psi(\theta_1)$. Si dependieran también de θ_2 , quiere decir que no nos podremos abstraer de la marginal de x_t para poder evaluar los parámetros de interés.
- **Exogeneidad fuerte:** Este caso se usa para hacer proyecciones de y_t , las cuáles se alimentan de proyecciones de x_t . En este caso, se deben cumplir dos condiciones:
 1. Que la variable x_t sea WE para ψ .
 2. y_t no cause en el sentido de Granger a x_t , lo que quiere decir que la marginal de x_t no depende de valores pasados de y_t o bien, para proyectar valores de x_t no son necesarios valores de y_t :

$$f(x_t|Y_{t-1}, X_{t-1}, \theta_2) = f(x_t|X_{t-1}, \theta_2)$$

- **Super exogeneidad:** No es algo que se requiera siempre. Si crítica de Lucas es una categoría científica y no metafísica, entonces debe ser falseable, lo cuál se verifica con superexogeneidad. Se deben cumplir dos condiciones:
 1. x_t WE para ψ .
 2. $f(y_t|x_t)$ es estructuralmente invariante: ante cambios en la marginal de x_t la condicional debe mantenerse estable (caso en el que se rechaza la crítica de Lucas).

2.5 Modelo Congruente

: Ahora, veremos los 6 aspectos para que un modelo sea congruente: Pasado Relativo, Presente Relativo, Futuro Relativo, Consistencia con teoría económica, admisibilidad en los datos y modelos rivales.

1. **Pasado Relativo:** Aspecto que todo modelo econométrico debería satisfacer. Dice si hemos sido capaces o no de usar eficientemente toda la información disponible hasta el momento. Lo que se debe cumplir es que los residuos del modelo deben ser **ruido blanco**. De otra forma, habría una parte sistemática y predecible en el ruido blanco, y por tanto no se estaría usando toda la información de manera eficiente.

Las formas de testear ruido blanco son las siguientes:

- **Durbin Watson:**

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2}$$

Este test la verdad es bastante inútil, pero servirá más adelante. Limitaciones: Hipótesis Alternativa es que los procesos siguen un proceso AR(1), lo que limita bastante el análisis. Además no es robusto a cuando se ponen rezagos de la variable dependiente.

- **Breusch-Godfrey:** Regresionar \hat{u}_t contra x_t y $\hat{u}_{t-1}, \dots, \hat{u}_{t-m}$. Note que por construcción $\text{corr}(\hat{u}_t, x_t) = 0$. Se deben ver para distintos valores de m . Finalmente, sabemos que $TR^2 \sim \chi_m^2$, por lo que con eso cerramos el test. Además, notar que este método sirve para cualquier proceso ARMA que sigan los errores. Además, como sabemos que bajo ciertas condiciones un MA puede ser aproximado por un AR, tan solo con la parte AR está bien. m no tiene por qué ser muy grande, pero se debe testear igual.

- **Test Q (Box-Pierce, Ljung-Box):** Nula: Ruido Blanco, Alternativa: no existe ningún tipo de autocorrelación, donde el test además es robusto a que el proceso siga un MA , AR o combinaciones.

En primer lugar, se deben calcular las autocorrelaciones de orden s :

$$r_s = \frac{\sum_{t=s+1}^T \hat{u}_t \hat{u}_{t-s}}{\sum_{t=1}^T \hat{u}_t^2}$$

Luego, evaluar la hipótesis conjunta de que todas las autocorrelaciones de orden s a orden m son estadísticamente iguales a 0 o no, donde la distribución del estadístico es una χ_m^2 :

$$Q_{BP} = T \sum_{s=1}^m r_s^2 \rightarrow \chi_m^2$$

$$Q_{BP} = T(T+2) \sum_{s=1}^m \frac{r_s^2}{T-s} \rightarrow \chi_m^2$$

En conclusión, se debe asegurar que el modelo tiene residuos que cumplen con ser ruido blanco en el sentido vectorial.

Otra cosa que se les podría pedir a los errores es que sean **innovación**: No debe haber nada que esté disponible en el set de información que tengamos y que no hayamos utilizado que nos sirva para explicar \hat{u}_t . Lo anterior puede ocurrir cuando se marginalizan ciertas variables.

Para testear lo anterior, se debe regresionar los \hat{u}_t contra las variables usadas en el modelo x_t , y el resto del set de información z_t . Después tomamos el R^2 de la regresión tal que hacemos un test de multiplicadores de lagrange:

$$TR^2 \rightarrow \chi_m^2$$

Donde m sería la cantidad de variables excluidas z_t . La hipótesis nula es que los z_t no pertenecen al modelo, por lo que si se rechaza quiere decir que los \hat{u}_t no pueden ser innovación, pues hay info relevante para predecir y_t en z_t .

Otro aspecto, es que los errores sean homocedásticos, lo cuál facilita la inferencia. Si no, hay que utilizar la matriz corregida de White.

Otro aspecto, es normalidad de los errores. Esto sirve para que no haya muchas diferencias entre distribuciones en muestras finitas y distribuciones asintóticas. Los test que se hacen para la normalidad son

- * **Test Jarque-Bera:** Toma en cuenta lo siguiente: Si la distribución es normal, media y varianza son suficientes para caracterizar todas las distribuciones de todas las series. Si normal, tercer momento $Skewness = 0$ y el cuarto momento $kurtosis = 3$. El test JB evalúa la conjunta de si la simetría es igual a 0 y la curtosis igual a 3:

$$JB = \frac{T}{6} \left[S^2 + \frac{(K-3)^2}{4} \right] \rightarrow \chi_2^2$$

Donde tenemos que

$$S = \frac{1}{T} \sum_{t=1}^T \left(\frac{z_t - \bar{z}}{s_z} \right)^3 = \frac{1}{T} \sum_{t=1}^T \left(\frac{\tilde{z}_t}{s_z} \right)^3$$

$S > 0$ implica que la tiene una cola derecha larga y $S < 0$ tiene la cola izquierda larga.

$$S = \frac{1}{T} \sum_{t=1}^T \left(\frac{z_t - \bar{z}}{s_z} \right)^4 = \frac{1}{T} \sum_{t=1}^T \left(\frac{\tilde{z}_t}{s_z} \right)^4$$

$K > 3$ tiene un pick más pronunciado (leptocúrtica), $K < 3$ es más plana (platicúrtica) en relación a la normal.

Si queremos testear uno de las dos, solamente nos olvidamos de una de ellas en el estadístico, y por tanto este tendería a una distribución χ_1^2 .

2. **Presente Relativo:** Evaluar si hemos hecho condicionamiento válido de variables contemporáneas a nuestro modelo. Si nuestro modelo no tiene variables contemporáneas, no debemos testear presente relativo. Supongamos el siguiente modelo:

$$y_t = \beta x_t + u_t$$

$$x_t = \delta z_t + v_t$$

Si queremos evaluar la ecuación 1 sin considerar la ecuación 2 (marginal de x_t), entonces debemos testear exogeneidad débil de x_t respecto del vector de parámetros de interés β . La manera de verificar lo anterior, es viendo que los errores de la condicional de $y_t|x_t$, u_t son ortogonales a los errores de la marginal de x_t , v_t .

Lo anterior se da dado que si $E(x_t, u_t) \neq 0$, entonces $\hat{\beta}$ será inconsistente. Así supongamos que $E(u_t, v_t) \neq 0$. Si queremos evaluar $\frac{\partial y_t}{\partial x_t}$, si el cambio en x_t se dio por v_t , entonces tenemos que $\frac{\partial y_t}{\partial v_t} = \frac{\partial y_t}{\partial x_t} \frac{\partial x_t}{\partial v_t} + \frac{\partial u_t}{\partial v_t} \neq \frac{\partial y_t}{\partial x_t} \frac{\partial x_t}{\partial v_t}$, y luego el efecto de x_t sobre y_t no será solamente β . Así, testaremos indirectamente si x_t es WE a β :

- **Test de Hausman:** Primero, estima un modelo para la marginal de x_t (primera etapa IV), y compute residuos \hat{v}_t . Puede ser un buen modelo AR que cumpla con criterio de pasado relativo. Luego, computamos una regresión de y_t contra x_t y \hat{v}_t , y finalmente testear si los \hat{v}_t son estadísticamente significativos.
- **Forma alternativa:** Estimar la marginal de x_t , y obtener residuos \hat{v}_t . Luego, regresionar y_t contra x_t y computar residuos \hat{u}_t . Finalmente, regresionar \hat{u}_t contra x_t y \hat{v}_t , y verificar si \hat{v}_t es estadísticamente significativo.

Estamos tomando en cuenta que los parámetros de la condicional y la marginal sean libres de variación, y que los parámetros de interés sean solamente una función de la condicional. Para escoger la marginal, no es necesario tener la mejor forma de x_t , sino que con una buena basta.

3. **Futuro Relativo:** Buscamos evaluar maneras de testear la crítica de Lucas: Cambios en la ley de movimiento de una variable de estado, debiese cambiar la manera en la que se relaciona la condicional con la marginal. Como testear:

- **Test Chow Secuencial:** OLS pre y post quiebre. Luego, evaluar si parámetros se mantienen antes y después del quiebre. Dummy para quiebre, y se puede hacer en nivel más interacciones. Sin embargo, esto tiene un problema: H_0 no hay quiebre estructural, versus H_1 de que hay quiebre estructural, pero fue en $t = T^*$. Se impone hipótesis de que conocemos cuál fue el período del quiebre. Si no conocemos, no se puede identificar. Por tanto, se puede hacer el test de manera secuencial, variando la nula. Sin embargo, hay mejores alternativas.
- **Test Ramsey:** No linealidades omitidas. Se estima modelo lineal $y_t = x_t \beta + u_t$. Se calcula \hat{y}_t , y finalmente se estima $y_t = x_t \beta + \dots + \beta_n \hat{y}_t^n + \varepsilon_t$. Trata de aproximar no linealidades omitidas. Sirve para evaluar inestabilidad de parámetros.

- **Test CUSUM:** Modelo para y_t y se busca evaluar si los parámetros son estables en la especificación. Se toman valores dichos de y_t con x_t , utilizando $\hat{\beta}_{t-1}$ es decir, usando información hasta $t - 1$ (one step forecast). Si estimadores de $\hat{\beta}_t$ son relativamente estables, $\hat{\beta}$ debieran ser más o menos constantes. Luego, el error de predicción de 1 período adelante no debería ser sustancialmente diferente de 0:

$$w_t = \frac{y_t - x_t' \beta_{t-1}}{\sqrt{1 + x_t (X_{t-1}' X_{t-1})^{-1} x_t}}, \quad W_t = \frac{1}{\sigma} \sum_{k=1}^t w_j \quad t = k + 1, \dots, T, \quad E(W_t) = 0$$

Donde σ la desviación estándar estimada de la muestra total. Se puede evaluar si el Out of Sample Forecast es $\neq 0$ o no, de manera gráfica utilizando la secuencia de W_t . Si la secuencia W_t no rechaza la nula, entonces los parámetros son estables. Si se rechaza la nula, podemos concluir inestabilidad de parámetros por lo que el modelo no sería útil para caracterizar proyecciones.

- **CUSUM of squares of residuals.** Mismos w_t de antes, pero elevados al cuadrado:

$$S_t = \frac{\sum_{j=t+k}^t w_j^2}{\sum_{j=t+k}^T w_j^2}, \quad E(S_t) = \frac{t - k}{T - k}$$

Se suma hasta t en el numerador, y se suma hasta T en el denominador. Cuanto porcentaje del total de la suma acumulada de los cuadrados se debe a los primeros t términos. Suma acumulada irá hasta 1: La expresión converge hasta 1. Se puede construir un intervalo de confianza, que serían dos líneas paralelas. Si suma está dentro del intervalo, todo ok. Si no, hay evidencia para rechazar la nula de estabilidad de parámetros. Tipicamente se hacen los dos CUSUM.

- **Recursive Least Squares:** β_t usa información de β_{t-1} para estimación. Por eso es recursivo. Idea básica: Estimar la estabilidad de parámetros en particular. Trayectoria del estimador con 10-100 datos. Se puede construir un intervalo de confianza asociado a ese estimador. Más datos implica menor varianza, y menos datos implica mayor varianza. Ver si el valor final de toda la muestra está o no comprometida en las intervalos de confianza que se tenía hacia atrás. Se rechazaría la nula de estabilidad de parámetros si el valor final no está siempre dentro del IC. Más exigente, ver si el IC final está siempre dentro del IC durante la muestra.
- **Evaluación fuera de muestra** Modelo confiable para predicción de escenario contrafactual, es necesaria la super exogeneidad: que la condicional de y_t sea estructuralmente invariante. Supongamos que $y_t = x_t' \beta + u_t$. Luego, $\frac{\partial y_t}{\partial x_t} = \beta$ si x_t es superexógena para β . Si se hace una intervención en x_t tal que cambia la manera en la que esta se genera, por crítica de Lucas, si marginal de x_t cambia, la condicional de x_t también debería cambiar.

Para ver que x_t sea super exógena a β , se necesita que ocurran dos cosas.

- i. Marginal de x_t debe ser inestable (se puede utilizar test CUSUM o CUSUM2).
- ii. Condicional de $y_t | x_t$ no cambia.

Si no se dan esas condiciones, no es posible testear si la Crítica de Lucas es o no válida.

4. **Teoría Económica:** Se debe cumplir que el modelo tenga sentido económico, y por tanto que cumpla con la interacción dinámica entre economía y evidencia empírica. Luego, en particular los signos y magnitudes de las estimaciones deben tener sentido, ya que el modelo económico es diferente del modelo estadístico.
5. **Admisibilidad en los Datos:** Este es un problema de medida. Resultados deben ser consistentes con propiedades que debiesen tener los datos. El modelo debe cumplir con las restricciones inherentes de los datos: Gini entre 0 y 1, tasa de desempleo positiva, etc.

6. **Modelos Rivales:** Hay modelos pasados sobre los cuáles se puede trabajar. Tratar de ver como se compara nuestro modelo con modelos alternativos, y ver si en el margen, nuestro modelo logra proveer información que los demás no. Los tipos de encompasamiento son los siguientes:

- **Varianza:** Si el modelo A tiene más varianza explicada que el modelo B, entonces el modelo A encompasa al modelo B. Este es un criterio destructivo, pues además no es consistente a la hora de seleccionar modelos. Modelos más grandes siempre tendrán mayor R^2 por lo que siempre encompasarían.
- Especificación
- Parámetros
- Misspecification
- Selección
- Parsimonioso: Se puede escoger un modelo que sea más simple, pero que explique resultados de un modelo más general. Además, este criterio cumple que si A encompasa a B, y B a C, entonces A encompasa a C (Transitividad).

En estas cosas, se debe notar que un modelo econométrico debería siempre encompasar a un modelo estadístico, dado que se tiene la ventaja de saber relaciones teóricas que los estadísticos no saben.

Si tenemos un modelo que cumpla con todos estos aspectos, tendremos un modelo congruente. Sin embargo, no es necesario que todos los modelos cumplan con estas condiciones, sino que depende de lo que se quiera hacer.

2.6 GETs en la práctica

Se comienza con un modelo sobreparametrizado, que debe cumplir con que los errores sean ruido blanco. Y desde ese punto, se comienza a podar el modelo mediante una reducción secuencial y testeo. Se puede eliminar los p-values mayores, y reduciendo su valor en la medida que aumentan las eliminaciones. Después de cada eliminación, testear si el modelo sobreviviente cumple con ruido blanco. Si cumple, seguir reduciendo hasta el mínimo posible. Luego, con el modelo podado, ver cuáles son las variables que se podrían asociar dado que tienen parámetros de valor similar entre ellas, para poder hacer un modelo más parsimonioso. Esto se puede hacer mediante una nula conjunta.

2.7 Correlaciones Espurias

Supongamos que :

$$y_t = t_{t-1} + u_t$$

$$z_t = z_{t-1} + v_t$$

Además, asumimos que $E(u_t, v_t) = 0$. Supongamos que tiramos una regresión entre ellos:

$$y_t = \delta z_t + w_t$$

En principio, como no hay relación entre las variables, los coeficientes no deberían ser significativos. Sin embargo, en gran parte de los casos el coeficiente es estadísticamente diferente de 0. De hecho, $Pr(|t| > 2) = 0.78$. Lo anterior se conoce como una correlación espuria. Ahora, veremos una manera en la que se puede desenmascarar una correlación espuria:

Si $R^2 < DW$, entonces lo más probable es que sea una relación espuria. Cuando se tienen dos variables $I(1)$, y se regresionan, el $DW \rightarrow 0$. Esto quiere decir que el coeficiente asociado a la autocorrelación de los residuos debería tender a 1, donde debiese persistir una raíz unitaria.

2.8 Cointegración

Comenzamos por lo siguiente. Cuando se tiene una serie $I(d)$ y otra $I(j)$, en general, cualquier combinación lineal de estas dos series será igual al mayor entre d, j . Luego:

$$I(d) \oplus I(j) = I(d), \quad d \geq j$$

Sin embargo, hay algunos casos en los que

$$I_X(d) \oplus I_Y(d) = I(d-1)$$

Donde decimos que X e Y cointegran. Además, no sirve cualquier combinación lineal. Solamente hay una única combinación que funciona, por lo que si:

$$Y_{I(1)} + \beta X_{I(1)} = Z_{I(0)}$$

Entonces solamente existe un único β sujeto a reparametrizaciones. En el ejemplo anterior, como las series no cointegran, entonces el DW de los residuos tiende a 0, pues estos son $I(1)$. En este caso, β es super consistente ($\mathcal{O}_p(T^{-1})$), y la inferencia se hace de manera estándar. Finalmente, la cointegración no quiere decir que haya una única dirección de la causalidad: $Y = f(X)$, $X = F(Y)$ pueden ser válidos. Solo se indica que ambas variables se mueven en conjunto en el largo plazo.

Para testear cointegración: H_0 es que no hay integración, y H_1 es que las series cointegran:

- Regresionar Y en X y luego hacer un test de DF o ADF en los residuos, y se toman los valores críticos de Phillips y Oullaris, dado que se debe tomar en cuenta que se hizo una regresión intermedia.
- Se puede hacer el DW y revisar que sobrepase unos valores críticos, según Engle y Granger.

2.9 Modelo de Corrección de Errores

Comenzamos con un modelo como el que sigue:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t, \quad y_t \sim iid(0, \sigma^2)$$

Consideramos ahora lo que se llama como relación de largo plazo para ambas variables, donde $c_t^* = c_t = c_{t-1}$:

$$y_t^* = \kappa_0 + \kappa_1 x_t^* \quad \kappa_1 = \frac{\beta_1 + \beta_2}{1 - \beta_3} \quad \kappa_0 = \frac{\beta_0}{1 - \beta_2}$$

Esto puede servirnos para testear teorías que modelan relaciones de largo plazo, como Friedman y la inflación como fenómeno monetario. Así, nos gustaría tener un modelo que ancle la teoría en el largo plazo, pero que sea capaz de entregarnos una buena representación de los datos en el corto plazo. Así, se puede re-escribir el modelo de arriba para llegar a la siguiente ecuación:

$$\Delta y_t = \beta_1 \Delta x_t + (\beta_2 - 1)(y_{t-1} - \kappa_0 - \kappa_1 x_{t-1}) + u_t \quad 0 < \beta_2 < 1$$

Donde hemos tenido el modelo de corrección de errores. El primer termino se llama el efecto impacto, el cuál plantea que el cambio que sufre y_t en un período dependerá de como se afecta un cambio en x_t . Lo que está entre paréntesis se llama término de corrección de errores: Esta plantea qué tan alejada está y_{t-1} de sus fundamentales de largo plazo consistentes con x_{t-1} . Aquí, pueden ocurrir dos cosas: Si $y_{t-1} > \kappa_0 + \kappa_1 x_{t-1}$, entonces como $1 - \beta_2$ es negativo, entonces todo el término es negativo y por tanto y_t debería ir cayendo, o bien, una corrección de error. Si tenemos que $y_{t-1} < \kappa_0 + \kappa_1 x_{t-1}$, entonces por el mismo argumento anterior se esperaría que y_t aumentara dado que hay un mecanismo natural que hace que y_t se acerque a su referente de largo plazo. Finalmente, $1 - \beta_2$ es el término de feedback que

nos dirá que tan rápido será el ajuste respecto de ajustes de largo plazo. Si β_2 se acerca a 0, entonces habrá una corrección más rápida, pues habrá menos persistencia en los shocks. Si β_2 se acerca a 1, hay más persistencia en los shocks, y por tanto, el ajuste demora más.

La especificación anterior es tremendamente útil:

- En primer lugar, anida a modelos que pueden ser en nivel o en diferencia.
- Es capaz de entregarnos información sobre la dinámica de largo plazo que vienen desde la teoría, y la estructura dinámica más flexible del corto plazo.
- El modelo es derivable desde los primeros fundamentos: Se podría derivar como si viniera de un proceso de optimización dinámico.

Dos maneras en las que se puede estimar el modelo

- Estimación por dos etapas: En la primera etapa evaluamos si y_t es $I(1)$, y también revisamos si x_t es $I(1)$ (o al menos, alguna de las variables x_t). Luego, se estima la ecuación de largo plazo por OLS ($y_t^* = \kappa_0 + \kappa_1 x_t^*$), recordando que los parámetros estimados serán super consistentes. Luego, se puede testear cointegración.
- Estimación directa la cuál se puede hacer mediante NLLS.

2.10 Sistemas Cointegrados

Cuando buscamos escribir el análogo al modelo de corrección de errores para un VAR, tenemos un Vector Error Correction. Idea es que se tendrá un $VAR(p)$, que puede ser representado en diferencias de la siguiente manera:

$$\hat{Y}_t = \alpha + \phi_1 \hat{Y}_{t-1} + \dots + \phi_p \hat{Y}_{t-p} + \phi_0 Y_{t-1} + u_t$$

Ahora, asumimos que cada y_t , i es $I(1)$, pero que h combinaciones lineales de Y_t son estacionarias, por lo que las combinaciones son $I(0)$, donde los errores son ruido blanco. h debe ser menor que el número de relaciones de largo plazo. Luego, la meta sería estimar todos los parámetros sujeto a

$$\phi_0 = -BA'$$

donde $Bh \times h$ y $A_{h \times n}$. Notar que lo anterior quiere decir que $Z_{t-1} = A'Y_{t-1}$ es $I(0)$, con lo que estaríamos encontrando vectores de cointegración (h) que permiten que las relaciones de largo plazo de Y_t estén ancladas. Para realizar la estimación, debemos utilizar el procedimiento de Johansen. Se busca estimar el modelo de arriba, pero identificando los vectores de cointegración. Cuantos y cuales son. Para lo anterior, se prosigue en dos pasos:

- **Se estima dos regresiones auxiliares:** un modelo donde la variable dependiente es la primera diferencia regresionada contra $p - 1$ rezagos del vector Y_t :

$$\hat{Y}_t = \alpha + \eta_1 \hat{Y}_{t-1} + \dots + \eta_p \hat{Y}_{t-p+1} + u_t$$

Luego, se hace una regresión similar, pero se pone como variable dependiente el nivel de Y_{t-1} en lugar de la diferencia:

$$Y_{t-1} = \alpha + \tau_1 \hat{Y}_{t-1} + \dots + \tau_p \hat{Y}_{t-p+1} + u_t$$

- **Se calculan las correlaciones canónicas**, la que sería más o menos la matriz de varianza y covarianza de los errores de cada regresión, y la covarianza de los errores de cada regresión.

$$\hat{\Sigma}_{ww} = T^{-1} \sum \hat{w}_t \hat{w}_t'$$

$$\hat{\Sigma}_{vv} = T^{-1} \sum \hat{v}_t \hat{v}_t'$$

$$\hat{\Sigma}_{vw} = T^{-1} \sum \hat{v}_t \hat{w}_t'$$

$$\hat{\Sigma}_{wv} = T^{-1} \sum \hat{w}_t \hat{v}_t'$$

- **Calcular valores propios.** Se debe calcular los eigenvalues (n en total) de la siguiente matriz:

$$\hat{\Sigma}_{ww}^{-1} \hat{\Sigma}_{wv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vw}$$

- Luego, sabemos que el máximo valor de MLE VEC sujeto a la restricción es el siguiente:

$$L^* = -(Tn/2) \ln(2\pi) - (Tn/2) - (T/2) \ln |\hat{\Sigma}_{vv}| + (T/2) \sum_{i=1}^h \ln(1 - \lambda_i)$$

Donde la gracia de lo anterior es que en el último término tenemos la suma de los primeros eigenvalues hasta el valor h, por lo que se puede testear que hay 1 relación de cointegración con $h = 1$, y así con k relaciones, testeando $h = k$. Finalmente, se puede hacer un test de razón de verosimilitud (LRT) para evaluar cuántas relaciones se tienen. Cuando se definen los h, faltarían los vectores de cointegración, los que serían los eigenvectores asociados a los valores propios.

3 Tópicos

3.1 ARCH

No todas las series tienen errores que distribuyen normal. Incluso, hay series que tienen una volatilidad que va cambiando, y no es constante, generando algutinamiento de volatilidades. Además, muchas series presentan colas anchas, lo que es conocido como leptocurtosis. Luego, la probabilidad de tener un boom o un crash son bastante mayores que las que típicamente se usan cuando se asume que los errores distribuyen normal. Ahora bien no siempre es malo asumir normalidad. Si tenemos un modelo como el que sigue:

$$y_t = \beta x_t + u_t, \quad u_t \sim (0, \sigma_t^2)$$

$$\sigma_t^2 = f(z_t, \alpha)$$

Luego, por FGLS, se puede tener un estimador consistente de β , y se puede utilizar una matriz robusta a heterocedasticidad para poder hacer inferencia válida, aún desconociendo el patrón de la heterocedasticidad. Ahora bien, en algunos casos puede resultar natural tener un estimador consistente de la varianza de esos procesos, en particular cuando no se cumple el supuesto de iid en la varianza de los errores y por tanto no se puede utilizar FGLS.

Un caso particular de la función $f()$, es cuando la volatilidad presenta dependencia temporal, y la volatilidad sigue un proceso análogo al de AR. Así, podemos caracterizar la volatilidad como un proceso autoregresivo $ARCH(m)$, Auto-Regressive-Conditional-Heteroskedasticity. Esto es, los cuadrados de los residuos de la regresión de y_t en x_t siguen un patrón de dependencia temporal:

$$u_t^2 = \varsigma + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2 + \omega_t$$

Donde m es la cantidad de rezagos utilizados en el modelo. Notar que los residuos cuadrados no dependen determinísticamente de sus residuos, sino que se incorpora un shock. Dado que tenemos formas para los dos primeros momentos, podemos calcular la esperanza de los u_t^2 :

$$\mathbb{E}(u_t^2 | u_{t-1}^2) = \sigma_t^2 = \varsigma + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2$$

Esta esperanza nos entregará información sobre la varianza condicional de los u_t , por lo que básicamente estamos diciendo que la varianza depende de los m rezagos de u_t^2 y no de la varianza rezagada. Así, la varianza es una función determinística de los u_t rezagados, con el problema de que u_t es una variable no observable. Notar además, que para que el proceso sea estacionario necesitamos que los eigenvalues del proceso sean menores que uno en modulo ($\sum \alpha_j < 1$). Y si así fuese, entonces la varianza incondicional del proceso u_t será igual a

$$h = \frac{\varsigma}{1 - \sum \alpha_j}$$

Otra forma de mirar el proceso, sería definiendo $u_t = \sqrt{h_t} v_t$ con $v_t \sim (0, 1)$ donde v_t es una serie homocedástica, multiplicada por una variable que va cambiando dependiendo del u_t , que sería la desviación estándar condicional de la serie:

$$h_t = \varsigma + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2$$

donde h_t sería la varianza condicional de u_t . Notar que para que sea una varianza, se requiere que todos los coeficientes de la regresión sean no negativos, ya que si no, en caso de shocks negativos muy fuertes, se tendría una varianza que es negativa lo cuál no tiene sentido:

$$\varsigma, \alpha_i \geq 0 \forall i$$

3.2 GARCH

Este es un análogo a un proceso ARMA pero para la varianza condicional. Luego, ahora los u_t^2 dependerá de m rezagos de sí misma, pero además se le agregarán rezagos de la varianza condicional h_t . Así, un proceso $GARCH(n, m)$ (Generalized ARCH) donde n son los rezagos de la varianza condicional y m son los rezagos de u_t^2 se vería de la siguiente manera:

$$u_t^2 = \varsigma + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2 + \delta_1 h_{t-1} + \dots + \delta_n h_{t-n} + \omega_t$$

$$\mathbb{E}(u_t^2 | u_{t-1}^2) = \sigma_t^2 = \varsigma + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2 + \delta_1 h_{t-1} + \dots + \delta_n h_{t-n}$$

Así, para saber si hay estacionariedad, debemos centrarnos en la parte AR del proceso (las que tienen u_t^2). Ahora, debemos imponer no negatividad a los parámetros δ :

$$\varsigma, \alpha_i, \delta_i \geq 0 \forall i$$

Finalmente, la varianza incondicional debe tomar en cuenta la parte ARCH y la parte GARCH, por lo que sería la siguiente:

$$h = \frac{\varsigma}{1 - \sum \alpha_j - \sum \delta_j}$$

3.3 Testeo

La H_0 de estos test es que hay homocedasticidad, mientras que H_1 es que hay $ARCH(m)$. Si bien hay otros test, nosotros nos centraremos en los que se testean por dependencia temporal en la volatilidad de las series. Luego, bajo la nula creemos que la varianza de los u_t es simplemente una constante, por lo que $\alpha_i, \delta_i = 0$. Esto sería lo mismo que testear que la serie sea ruido blanco, vs que no lo sea y que siga una dependencia temporal. Así, nos gustaría testear lo mismo pero en el segundo momento. Si bien no observamos u_t como un β_{OLS} sí es consistente, podemos computar de manera consistente los u_t . Y finalmente, podemos utilizar cualquier test sobre la serie resultante para ver si efectivamente los errores son ruido blanco o no (Ljung Box). Note que la nula se evalúa en el segundo momento. Luego Ljung Box tiene la gracia de que la alternativa es bastante general, por lo que el proceso que siguen los errores podría ser un GARCH cualquiera.

3.4 Estimación

Note que no se puede estimar por FGLS por cuanto la varianza condicional depende de los u_t^2 rezagados, lo cuál puede ser escrito como $y_t - \beta x_t$. Así, FGLS no será una manera eficiente de estimar, ya que se debe hacer en dos etapas. Lo más eficiente sería estimar por MLE. Así, asumiendo que u_t distribuye normal, tenemos

$$y_t = x_t' \beta + u_t$$

$$f(y_t | x_t) = \frac{1}{\sqrt{2\pi}h_t} \exp\left(-\frac{(y_t - x_t' \beta)^2}{2h_t}\right)$$

$$h_t = \zeta + \alpha_1(y_{t-1} - x_{t-1}' \beta)^2 + \dots + \alpha_m(y_{t-m} - x_{t-m}' \beta)^2 + \delta_1 h_{t-1} + \dots + \delta_n h_{t-n}$$

Por lo que la función de log-verosimilitud tendría la siguiente forma:

$$L = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=m+1}^T \ln h_t - \frac{1}{2} \sum_{t=m+1}^T \frac{(y_t - x_t' \beta)^2}{h_t}$$

Así, la estimación se debe realizar con métodos numéricos, imponiendo restricciones de estacionariedad y no negatividad. Notar que se debe comenzar con un guess para los primeros h , para que después estos mismos se vayan creando dependiendo de su dinámica.

3.5 Enfoques Alternativos

Aquí, veremos un par de modelos que cumplen con diferentes características:

- **Gallant y Tauchen:** Prononen estimar la desviación estándar condicional, lo cuál se hace de la siguiente manera: $\sqrt{h_t} = \zeta + \alpha_1 |IGARCH : Integrated GARCH. En este caso, se asume que $\sum \alpha_j + \sum \delta_j = 1$ y por tanto, la varianza tiene una raíz unitaria.$

•

- **TGARCH:** Threshold GARCH. Hay un cierto nivel de asimetría. Por ejemplo en períodos de shocks positivos, hay mayor volatilidad que en los shocks negativos. Así, se podría estimar un

$$h_t = \zeta + \alpha_1 u_t^2 + \delta \phi_{u_t > 0}$$

Así, la varianza dependerá de si el modelo supera cierto umbral.

- **ARCH in mean:** La esperanza condicional de y_t depende además de la varianza h_t , por lo que veríamos una especie de retroalimentación entre el primer y segundo momento. Volatilidad puede afectar el primer momento:

$$y_t = \beta' x_t + \gamma h_t + u_t, \quad u_t \sim (0, h_t^2)$$

$$h_t = \zeta + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2$$

Lo anterior puede complicar las cosas a la hora de la estimación por la definición de u_t .

- **EGARCH:** Exponential GARCH. Propone una forma funcional para h_t : $h_t = e^{z_t' \delta}$. Esto garantiza que h_t sea siempre no negativo, por lo que no hay que poner restricciones de no negatividad. Problema: varianza depende exponencialmente del modelo, por lo que se debe guardar para ocasiones donde se cumplan los supuestos para usarlo.

3.6 Desvíos de Normalidad.

Leptocurtosis: ARCH y GARCH generan algo, pero no es suficiente. Esto sirve para poder estudiar eventos extremos, para lo cual es fundamental generar algo de Leptocurtosis. Un acercamiento típico y popular es utilizar u_t como una t-student con v grados de libertad:

$$f(u_t) = \frac{\Gamma(\frac{v+1}{2})}{(\pi v)^{1/2} \Gamma(\frac{v}{2})} M_t^{-\frac{1}{2}} \left[1 + \frac{u_t^2}{M_t v} \right]^{-\frac{v+1}{2}}$$

Donde $\Gamma(\cdot)$ es la función gamma. Si $v > 1$ entonces $\mathbb{E}(u_t^2) = \frac{M_t v}{v-2}$. Notar que si v tiende a infinito, tendríamos una normal. Los grados de libertad no necesariamente son enteros en este caso. Se podría hacer que el v dependiera de otra cosa, proceso para h y otro para v .

No linealidades: procesos asimétricos: no tan utilizados en series de macro-finanzas. Tienen su problema en el 4to momento. Aproximador universal de funciones de densidad de u_t : SNP. Garantiza que se podrá aproximar cualquier función con segundo momento acotado.

3.7 Mixture de Normales

Se puede intentar aproximar cualquier tipo de función de densidad relativamente suave, continua, utilizando combinaciones de normales. Luego esta combinación tomaría la siguiente forma:

$$p(y|\rho) = \sum_{j=1}^J \pi_j N(\mu_j, \sigma_j^2)$$

Donde tenemos ponderadores π_j y normales condiferentes medias y diferentes varianzas. Combinando normales, se puede aproximar a una distribución que no necesariamente viene de una normal. Así, tenemos que la serie estudiada con cierta probabilidad vendrá de una normal, y con otra probabilidad de otra, y así. Ahora, para definir de dónde vienen las normales, asumiremos que esta probabilidad está asociada a una variable latente (es decir, no observable) s , que sería una variable de estado o de régimen. Esta dice que con cierta probabilidad, $s = i$ quiere decir que la serie viene de la normal número i . Si se pudiese observar simultáneamente la variable y y además la variable s , se podría caracterizar la distribución conjunta de ambas variables, la que por teorema de Bayes sabemos que se puede representar como el producto entre la condicional y la marginal:

$$P(y_t, s_t = j) = \frac{\pi_j}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{y_t - \mu_j}{2\sigma_j^2}\right) = P(y_t|s_t = j)P(s_t = j)$$

Note que en este caso, los parámetros a estimar serían los μ_j, σ_j, π_j . Y para estimar esto mediante MLE, se debe escribir en función de observables y parámetros, pero recordemos que s es no observable por lo que no se puede escribir la función de densidad conjunta. Para eso, debemos obtener cuál es la distribución incondicional de y , la que sería la suma (pues s es discreta) de las densidades conjuntas:

$$f(y_t) = \sum_{j=1}^J P(y_t, s_t = j)$$

Con lo que podremos plantear la log-likelihood, que será bastante no lineal (implicando que se deben escoger algoritmos numéricos con cuidado, para llegar al máximo global en vez de local).

$$L(\theta) = \sum_{t=1}^T \log f(y_t)$$

Note que a la hora de estimar, se debe garantizar que $0 < \pi_j < 1$ y además que $\sum \pi_j = 1$. Así, conviene reparametrizar y escribir los π_j de la siguiente manera:

$$\pi_j = \frac{a_j^2}{1 + \sum_{j=1}^{J-1} a_j^2}$$

Así, se debe derivar respecto de a_j asegurando que todas las probabilidades estarán entre 0 y 1. Note además de que se deben estimar $J - 1$ parámetros, pues el J debe ser tal que la suma sea igual a 1:

$$\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$$

Algo que se podría querer saber, es dadas la observación y_t , cuál es la probabilidad de que venga de la normal j , lo cuál se puede obtener nuevamente por el teorema de Bayes:

$$P(s_t = j | y_t) = \frac{P(y_t, s_t = j)}{f(y_t)} = \frac{\pi_j \cdot f(y_t | s_t = j)}{f(y_t)}$$

Se podría tener estimadores de $P(s_t = j | y_t)$.

3.8 Cadenas de Markov

Esta es una contraparte discreta de los modelos AR, pero los estados tendrán realizaciones discretas. Una cadena de markov de n estados quiere decir que hay n posibles realizaciones de s . Así, cuando se tiene una variable discreta con dependencia temporal, se utiliza cadena de markov.

$$P(s_t = j | s_{t-1} = i, s_{t-2} = k_1, \dots) = P(s_t = j | s_{t-1} = i) = P_{ij}$$

Una cadena de markov de primer orden, quiere decir que lo que va a ocurrir en el período t , solamente depende de lo que ocurrió en $t - 1$. La matriz de transición es aquella que resume todas las probabilidades de transición:

$$P = \begin{bmatrix} P_{11} & \cdots & P_{N1} \\ \vdots & \ddots & \vdots \\ P_{1N} & \cdots & P_{NN} \end{bmatrix}$$

Con $\sum_{j=1}^N P_{ij} = 1$. Esta matriz lo que nos da es la persistencia de quedarnos en alguno de los estados, lo cuál dependerá de la probabilidad de mantenerse en el mismo. Note que si todas las probabilidades son iguales, entonces no habría persistencia. Finalmente un proceso de markov de orden k es decir

que la probabilidad de estar en un estado depende de los últimos k estados, este se puede representar como si fuera un proceso de markov de orden 1.

El análogo a estacionariedad, es la cadena de markov ergótica, que nos dice cuál es la probabilidad incondicional de ocurrencia de un evento $\pi_{n \times 1}$. Esta cadena se puede encontrar a través de P de dos maneras posibles:

La primera sería a través de los valores propios de la matriz P . Como todas las columnas de la matriz suman 1, siempre habrá un $\lambda_i = 1$. Así, debemos saber cuál es el eigenvector de ese eigenvalue, lo que nos dará lugar a la cadena de markov ergótica después de reescalarlo dado que no necesariamente la suma de ese eigenvector será igual a 1.

$$P\pi = \pi$$

$$\lim_{n \rightarrow \infty} P^n = \pi \mathbf{1}'$$

Otra forma de hacerlo sería encontrando la columna $n + 1$ de la matriz B :

$$B = (A'A)^{-1}A', \quad A = \begin{bmatrix} I_N - P \\ \mathbf{1}' \end{bmatrix}_{(N+1) \times N}$$

3.9 Markov Switching Regimes

Asumiremos que existen dos regímenes, pero es válido para n regímenes de todas formas. Variable que determina el régimen no es observable por el econometrista.

$$\mathbb{L} = \sum_{t=1}^T \ln[p_{11}f_1(y_t|x_t\theta_1) + (1 - p_{11})f_2(y_t|z_t\theta_2)] + [(1 - p_{22})f_1(y_t|x_t\theta_1) + p_{22}f_2(y_t|z_t\theta_2)]$$

Donde $f_1(\cdot)$, $f_2(\cdot)$ pueden ser cualquier modelo, no es necesario que sean el mismo ni que dependan de las mismas variables. Regímenes que ocurren no son permanentes. La transición de un régimen a otro, depende de un proceso de Markov, con la siguiente matriz de transición

$$P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}$$

Lo que ocurre ahora, es que los regímenes no son observables por el econometrista, sino que solamente vemos y_t , x_t . s es una variable latente. Buscar tratar de encontrar parámetros P , θ_1 , θ_2 solamente observando y_t , x_t . En este caso, como hay dependencia temporal, no se puede obtener la densidad de y_t , por lo que no se puede aplicar lo mismo que en MN. El algoritmo que se puede utilizar en estos casos es el Expectation Maximization (EM) algorithm. Básicamente se busca maximizar un valor esperado tomando en cuenta que no se observa el régimen por el que se comienza:

- Comenzamos con valores iniciales, y luego se trata de definir condicional en el dato que observamos y en los parámetros que partimos, cuál es la probabilidad de que en cada período estemos en un determinado régimen
- Con esas probabilidades calculadas, se puede retroalimentar el proceso obteniendo un nuevo guess, y así sucesivamente.

Hay extensiones posibles que se pueden hacer a estos modelos, como endogeneizar las probabilidades de ocurrencia dejándolas en función de observables (Lee y Nelson).

3.10 Modelos de umbral

Más convenientes que los modelos de markov switching: más fáciles de estimar, más fáciles de interpretar y más fáciles de especificar. Como la transición de régimen a régimen depende de variables observables, es más fácil dar una interpretación económica al modelo.

Veremos un modelo bastante simple, pero fácilmente generalizable. Supongamos que tenemos lo siguiente:

$$y_t = (x_t' \beta) I_{1_t \leq \gamma} + (z_t' \delta) I_{q_t > \gamma} + u_t$$

$$y_t = x_t'(\gamma) \theta + u_t, \quad \theta = (\beta', \gamma')'$$

Donde la variable y_t depende (no necesariamente linealmente) del conjunto de variables x_t , y de otro conjunto de variables z_t . Lo que quiere decir que y_t reacciona de manera diferente dependiendo de una función indicador, la que se gatilla cuando se cumple la condición que está en el subíndice, la que es observable por el econometrista. Luego, notar que hay un quiebre dependiendo de si q_t logra superar un cierto umbral. Si hay más de 2 regímenes, entonces también estará $\gamma_1 < q_t < \gamma_2$, con lo que se tendrían dos parámetros de umbral. Note que se debe estimar los parámetros γ , pero la función no es diferenciable en tales parámetros, pero se puede utilizar el método de concentración. Así, condicional en un valor de γ , se puede tener los estimadores de antes del quiebre y después del quiebre:

$$\hat{\theta}(\gamma) = \left(\sum_{t=1}^T x_t(\gamma) x_t(\gamma)' \right)^{-1} \left(\sum_{t=1}^T x_t(\gamma) y_t \right)$$

Así, una vez que se conoce γ estimar todo lo demás es trivial, pues es solo incluir un par de dummies. Notar que esto es una ventaja respecto de Markov Switching, ya que en ese caso nunca se podría hacer el sample splitting por cuanto no hay una variable que genere el quiebre. Ahora, para estimar el valor de γ , se debe hacer una grilla con todos sus posibles valores, y obtener los $\hat{u}_t(\gamma)$ de ese modelo. Con eso, se puede construir la varianza (suma del cuadrado de los residuos), y se busca a el estimador de γ que minimice tal suma, es decir, que haga lo más pequeño posible el componente no sistemático::

$$\hat{\sigma}^2(\gamma) = \frac{1}{T} \sum_{t=1}^T [\hat{u}_t(\gamma)]^2$$

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} \hat{\sigma}^2(\gamma), \quad \Gamma = [\underline{\gamma}, \bar{\gamma}]$$

$$\rightarrow \hat{\theta}(\hat{\gamma}) = \hat{\theta}, \quad \hat{\sigma}^2(\hat{\gamma}) = \hat{\sigma}^2$$

Ahora, veremos algunas consideraciones prácticas sobre como escoger la variable de umbral. Algunas veces esta viene dada por la teoría económica, mientras que en otros casos habría que encontrarlas.

En primer lugar, veremos el Self Exciting Threshold (SET). La variable q es la misma variable dependiente pero rezagada.

Otros modelos podrían pensar en q como el tiempo t , dando lugar a un modelo de quiebre estructural. En estos casos, se asume que el quiebre es para siempre, y que nunca se regresará al régimen anterior. Ahora, note que el econometrista no necesariamente conoce el período del quiebre, por lo que el mismo es endógeno y por tanto, se podría aplicar

Finalmente, en caso de no conocer el q , se puede probar con varias alternativas de variables y escoger según algún criterio de información consistente.

Para terminar, queremos probar efectivamente que hay un umbral según el cuál la dinámica de las variables cambia. Así, la nula que nos gustaría testear es si $\beta = \delta$. Ahora, como se puede testear?

Ahora, cuando $\beta = \delta$, tenemos que el parámetro γ no está identificado bajo la nula. Luego, los test convencionales para evaluar la nula no funcionarán:

$$F(\gamma) = T \left(\frac{\sigma^2 - \hat{\sigma}(\hat{\gamma})}{\hat{\sigma}^2(\hat{\gamma})} \right)$$

Donde lo de arriba sería una especie de variante de un test de Chow, solamente que no distribuye χ^2 dado que γ no está identificado. Así, lo que se debería hacer es un bootstrap y comparar de acuerdo a los valores críticos bajo la nula.

Otra forma, sería modificar el modelo para que en vez de ser una step function tenga una transición continua, donde pueden aparecer las funciones logísticas, etc.

4 Non Linear Least Squares

Decimos que una función de regresión $m(x_t, \beta) = \mathbb{E}(y_t | x_t, \beta)$ es no lineal en los parámetros si no se puede escribir como $m(x_t, \beta) = f(x_t)' \beta$ para alguna función $f(\cdot)$. Por ejemplo:

$$m(x, \beta) = \beta_1 + \beta_2 + \frac{x}{1 + \beta_3 x}$$

4.1 Estimación

El estimador se puede definir como el que resuelve el siguiente problema:

$$\hat{\beta}_{NLLS} = \arg_{\beta} \min S_T(\beta)$$

donde

$$S_T(\beta) = \sum_{t=1}^T (y_t - m(x_t, \beta))^2 = [Y - m(X, \beta)]' [Y - m(X, \beta)]$$

Definimos el jacobiano como los vectores de gradiente para cada observación:

$$Z(\beta)_{T \times k} = \frac{\partial m}{\partial \beta'} = \begin{bmatrix} \frac{\partial m_1}{\partial \beta_1} & \dots & \frac{\partial m_1}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial m_n}{\partial \beta_1} & \dots & \frac{\partial m_n}{\partial \beta_k} \end{bmatrix}$$

Luego, los parámetros de *NLLS* deben resolver las siguientes condiciones de primer orden:

$$\left. \frac{\partial S_T(\beta)}{\partial \beta} \right|_{\hat{\beta}} = -2Z(\hat{\beta})' [Y - m(X, \hat{\beta})] = 0$$

De lo anterior, note que OLS es un caso particular de NLLS. En segundo lugar, no se puede resolver la ecuación de arriba de manera analítica, por lo que se requerirán métodos numéricos. Por último, como la FOC son no lineales, pueden haber más de una raíz, por lo que se necesitará una evaluación para encontrar mínimos globales.

4.2 Métodos Numéricos: Gauss Newton

Considere una serie de Taylor de primer orden para $m(x_t, \beta)$ comenzando por el valor β_0 .

$$m(x_t, \beta) \approx m(x_t, \beta_0) + \left. \frac{\partial m_t}{\partial \beta'} \right|_{\beta_0} (\beta - \beta_0)$$

Reemplazando en la función objetivo, tenemos:

$$S_T(\beta) = \sum_{t=1} \left[y_t - m(x_t, \beta_0) - \frac{\partial m_t}{\partial \beta'} \Big|_{\beta=\beta_0} (\beta - \beta_0) \right]^2$$

Como esta función es cuadrática en β , las FOC serán lineales en el mismo parámetro y por tanto, habrá solución analítica. Resolviendo lo anterior, llegamos a:

$$\beta_1 = \beta_0 + [Z(\beta_0)'Z(\beta_0)]^{-1} [Z(\beta_0)(Y - m(X, \beta_0))]$$

Se repite el proceso hasta que se llegue a una convergencia ($\beta_j \approx \beta_{j-1}$) y definimos $\beta_j = \hat{\beta}_{NLLS}$. Con lo anterior, podemos derivar estimadores consistentes para σ^2 : $\hat{\sigma}^2 = T^{-1} S_T(\hat{\beta})$, o $\hat{\sigma}^2 = (T - k)^{-1} S_T(\hat{\beta})$

Note que también se podría pensar la estimación Gauss-Newton como una regresión lineal si es que se reescribe el modelo de la siguiente manera:

$$\begin{aligned} Y - m(X, \beta_0) + Z(\beta_0)\beta_0 &\approx Z(\beta_0)\beta + u \\ \rightarrow Y^* &\approx X^* + u \end{aligned}$$

Donde podemos llegar al mismo resultado del parámetro anterior. La ventaja del algoritmo Gauss-Newton es que es fácil de computar la iteración, y no requiere uso de Hessianos. Finalmente, el método solamente puede ser aplicado en problemas equivalentes a la minimización de la función objetivo de NLLS.

Como consideraciones prácticas, $Z(\beta_j)'Z(\beta_j)$ debe ser invertible para todo j , y en segundo lugar es útil obtener las derivaciones analíticas del jacobiano. Además, debido a la naturaleza no lineal del problema, pueden haber múltiples raíces, por lo que se recomienda partir con Guess distintos y luego evaluar en la función objetivo para ver cuál es menor. Un consejo útil sería graficar las funciones.

Finalmente hay dos posibles criterios de convergencia: Relativa y Absoluta. La primera mide el porcentaje de diferencia que hay entre una iteración y otra $tol > |\frac{\beta_j - \beta_{j-1}}{\beta_{j-1}}|$, mientras que la absoluta ve la diferencia absoluta que hay entre las dos iteraciones $tol > |\beta_j - \beta_{j-1}|$. Finalmente, cuando se cumple la condición de tolerancia se podría verificar si es que el estimador encontrado se aproxima a la solución de las FOC, es decir si $\frac{\partial S_T(\beta)}{\partial \beta} \Big|_{\beta_j}$ se aproxima a 0.

4.3 Métodos Numéricos: Concentración

Este método se puede lograr cuando se crea la siguiente partición: $\beta = (\gamma, \delta)$ tal que

$$m(x_t, \beta) = \gamma' x_t(\delta)$$

Donde $x_t(\gamma)$ es una función $k \times 1$ de x_t y de δ . Luego, se toma una grilla para los δ , y se hace la siguiente minimización secuencial:

$$\min_{\beta} S_T(\beta) = \min_{\delta} \min_{\gamma} S_T(\gamma, \delta)$$

Así, como el modelo es lineal en γ vemos que:

$$\hat{\gamma}(\delta) = \arg_{\gamma} \min S_T(\gamma, \delta)$$

$$\rightarrow = [X(\delta)'X(\delta)]^{-1} X(\delta)'Y$$

Finalmente, dejamos todo en función de δ :

$$S_T(\delta) = S_T(\hat{\gamma}(\delta), \delta)$$

Que sería la suma de residuos al cuadrado concentrada. Finalmente, tenemos que $\hat{\delta} = \arg_{\delta} \min S_T(\delta)$ por lo que $\hat{\gamma} = \hat{\gamma}(\hat{\delta})$. Finalmente, el par $(\hat{\gamma}, \hat{\delta})$ serán los estimadores NLLS de (γ, δ) . Los beneficios de la concentración son que la dimensión de la optimización numérica disminuyen mucho. Cuando δ es un escalar, se puede usar una grilla.

4.4 Inferencia con Restricciones Lineales.

No es difícil obtener una matriz de varianza covarianza:

$$V(\beta_{NLLS|X}) = \sigma^2 (X^{*'} X^*)^{-1}$$

o bien, definiendo $\hat{Z} = Z(\hat{\beta}_{NLLS})$ tenemos:

$$V(\beta_{NLLS|X}) = \sigma^2 (\hat{Z}' \hat{Z})^{-1}$$

Donde podemos utilizar el estimador de σ^2 desarrollado arriba, que bajo ciertas circunstancias, es consistente. Así, nos centraremos en las propiedades asintóticas más que de muestras pequeñas. La inferencia causal, se puede hacer normalmente.

4.5 Inferencia con Restricciones Lineales: test t

Es igual que en el caso de estimadores OLS:

$$\frac{Q' \hat{\beta} - c}{\hat{\sigma} [Q' (\hat{Z}' \hat{Z})^{-1} Q]^{1/2}} \sim^a S_{T-k}$$

donde S_{T-k} es la distribución t -student. Así, $T \rightarrow \infty$ tenemos que el test tiende a una normal $(0, 1)$. Crear los intervalos de confianza es lo mismo:

$$Pr \left[\hat{\beta}_i - z_{\alpha/2} \sqrt{\hat{V}_{i,i}} < \beta_i < \hat{\beta}_i + z_{\alpha/2} \sqrt{\hat{V}_{i,i}} \right]$$

4.6 Inferencia con Restricciones Lineales: test F

Test LRT:

$$\frac{S_T(\bar{\beta}) - S_T(\hat{\beta})}{\hat{\sigma}} \rightarrow \chi_q^2$$

Test Wald:

$$\frac{(Q' \hat{\beta} - c)' \left[Q' (\hat{Z}' \hat{Z})^{-1} Q \right] (Q' \hat{\beta} - c)}{\hat{\sigma}^2} \rightarrow \chi_q^2$$

4.7 Inferencia con Restricciones Lineales: Identificación

Pueden darse casos en los que bajo la H_0 , existan parámetros que no estén identificados. Por ejemplo:

$$m(x_t, \beta) = \beta_1' z_t + \beta_2' x_t(\delta)$$

Donde la nula es $H_0 : \beta_2 = 0$. En este caso los parámetros δ no estarían identificados. En estos casos, las distribuciones asintóticas son no estándar y por tanto, se requieren técnicas de simulación como bootstrap para construir intervalos de confianza.

5 GMM

5.1 Método de Momentos

Se busca hacer un match entre los momentos poblacionales y sus contrapartes muestrales. Así, tendremos un vector de vectores que dependen de los x_t $f(x_t, \theta)_{q \times 1}$, y sabemos que la esperanza condicional de la función debe existir y ser finita para todo t y θ . Así, la esperanza poblacional de la función evaluada en el verdadero vector de parámetros debe ser igual a 0:

$$\mathbb{E}(f(x_t, \theta_0)) = 0$$

Luego, en lugar de la esperanza condicional, utilizamos al promedio muestral, y buscamos el $\hat{\theta}$ que hace que el promedio muestral sea igual a lo que debiese ser la esperanza poblacional (0)

Método de momentos: caso en particular en el que tenemos que el número de parámetros a estimar es igual al número de momentos que queremos evaluar. Como ejemplo, supongamos que buscamos estimar la esperanza de x_t . Sabemos que

$$\mathbb{E}(x_t) = \theta_0 \rightarrow \mathbb{E}(x_t - \theta_0) = E(f(x_t, \theta_0)) = 0$$

Y luego, aproximamos la esperanza por el promedio muestral:

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^T x_t$$

Luego usamos la contraparte muestral de la esperanza poblacional como estimador, lo que es un estimador del método de momentos. Como ejemplo, se podría tomar el estimador de OLS, y de Variables Instrumentales. Para IV, tenemos que $\mathbb{E}(u_t | x_t) \neq 0$, pero considere un vector de q instrumentos tal que $\mathbb{E}(u_t | z_t) = 0$. Así, tenemos que lo anterior implica:

$$\mathbb{E}(z_t u_t) = \mathbb{E}(z_t (y_t - x_t \beta_0)) = 0$$

Por lo que usando la contraparte muestral, tenemos que el estimador de variables instrumentales es:

$$\frac{1}{T} \sum_{t=1}^T x_t (y_t - x_t' \hat{\beta}) = 0 \rightarrow \hat{\beta}_{SIV} = (Z'Z)^{-1} Z'Y$$

Que sería el estimador clásico por variables instrumentales.

5.2 Método Generalizado de Momentos

Este método es filosóficamente igual al de Momentos, solamente que los modelos no estarán exactamente identificados. Si tenemos más momentos (q) que parámetros a estimar (p), entonces tendríamos un modelo sobreidentificado, y por tanto, se necesita otra técnica de estimación. Si tenemos un sistema subidentificado ($q < p$), no se puede estimar.

En este caso, como tenemos más ecuaciones que incógnitas, tendremos infinitas soluciones. La idea ahora es la misma que teníamos antes: Sea $f_T(\theta) = T^{-1} \sum f(x_t, \theta)$ la contraparte muestral de los momentos poblacionales $f(\cdot)$ para cualquier vector de parámetros. Así, se debe encontrar un vector de parámetros tal que $f_T(\theta)$ se acerque lo más posible a 0. Note que como estamos en un modelo sobreidentificado, a diferencia del exactamente identificado, el vector de parámetros nunca logrará hacer que la suma sea igual a 0. Así, se crea una métrica de éxito mediante la minimización de una función objetivo:

$$Q_T(\theta) = f_T(\theta)'_{1 \times q} A_T{}_{q \times q} f_T(\theta)_{q \times 1}$$

Note que la función es cuadrática en f_T , por lo que es como minimizar una suma ponderada del cuadrado de los f_T , dado que A_T es una llamada matriz de ponderación (que es positiva definida), la cuál dice la importancia relativa que se le dará a cada momento para escoger el estimador, y además, la función Q_T es un escalar:

$$\hat{\theta}_T = \arg_{\theta} \min Q_T(\theta)$$

Por ejemplo, de variables instrumentales sobreidentificado es un caso particular de GMM. Ahora, veremos el ejemplo de por qué GMM es un método de estimación particularmente útil. Supongamos que tenemos un agente que maximiza su utilidad a lo largo de toda la vida sujeto a una restricción presupuestaria:

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

$$y_t + (1 + r_{t-1})b_{t-1} \geq c_t + b_t$$

Por lo que sabemos que la condición de Euler es (asumiendo forma funcional CRRA para $u(c_t)$):

$$\mathbb{E} \left[\beta \left(\frac{y_{t+1}}{y_t} \right)^{-\gamma} (1 + r_t) - 1 \right] = 0$$

De la ecuación de arriba, si se tuvieran datos sobre y_t y de r_t se podría querer estimar los parámetros β, γ , pero en ese caso se tendría más parámetros que momentos dado que se tiene solo una ecuación de Euler, y dos parámetros a estimar. Sin embargo, note que la esperanza aplicada es condicional a toda la información en t , por lo que debe ser ortogonal a cualquier cosa que se conozca en el período t , por lo que esa esperanza resume eficientemente toda la información en t . Luego, se puede utilizar como instrumento cualquier variable conocida en t para generar condiciones de momentos:

$$\mathbb{E} \left[\beta \left(\frac{y_{t+1}}{y_t} \right)^{-\gamma} (1 + r_t) - 1 \right] z_t = 0$$

Así, se podría poner cualquier cosa que se conozca en t , dado que la teoría es bastante fuerte en ese sentido.

5.3 Consistencia

Para saber si el estimador es consistente, se utilizarán las propiedades de estimadores extremos. Así, se deben cumplir los siguientes supuestos:

1. $\mathbb{E}(f(x_t, \theta))$ y es finito $\forall t, \theta$
2. $\mathbb{E}(f(x_t, \theta)) = 0$ si y sólo si $\theta = \theta_0$ Solo se cumple que la esperanza poblacional es igual a 0 solamente con θ_0 . Este es el supuesto de identificación.
3. $f_{T,j} = T^{-1} \sum f^j(x_t, \theta)$ y $g_{T,j} = T^{-1} \sum \mathbb{E} f^j(x_t, \theta)$ para $j = 1, \dots, q$ tal que $f_{T,j} - g_{T,j} \rightarrow p0$ uniformemente $\forall j, \theta$
4. $A_T - \bar{A}_T \rightarrow p0$ donde \bar{A}_T es una matriz no estocástica.

Si se cumple lo anterior, sabemos que $\hat{\theta} \rightarrow p \theta_0$

5.4 Normalidad Asintótica

Para obtener normalidad asintótica, además de los supuestos de consistencia debemos agregar un quinto:

$$5. \sqrt{T}f_T(\theta_0) \rightarrow dN(0, \bar{V}_T), \text{ donde } V_T = T \text{ Var}(f_T(\theta_0))$$

Si se cumple, entonces tenemos que

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N(C^{-1}BC^{-1})$$

$$C = F_T' A_T F_T \quad B = F_T' A_T \bar{V}_T A_T F_T, \quad F_T = \frac{\partial f_T}{\partial \sigma'}$$

Finalmente, para lograr eficiencia asintótica, debemos escoger A_T tal que converga en probabilidad a $\bar{A}_T = \bar{V}_T^{-1}$. Ahora bien, notar que si se escoge cualquier A_T el estimador de GMM sigue siendo consistente. Para escoger el A_T , se debe escoger un estimador consistente de la matriz de varianzas utilizando el $\hat{\theta}$.

5.5 Estimación

Note que en la parte anterior, para lograr obtener un estimador consistente de \bar{V}_T , se debía conocer un estimador consistente de θ_0 . Luego, estimar los dos al mismo tiempo podría parecer complicado. Así, debemos seguir el siguiente procedimiento de estimación:

En primer lugar, usamos un A_T que sea definida positiva y no estocástica (como por ejemplo, I_q lo cuál hace sentido pues no tenemos como saber cuál es el momento que más debería ponderar). Después, obtenemos los estimadores consistentes de GMM, obteniendo los estimadores de la primera ronda θ_I . Después, obtenemos un estimador consistente de \bar{V}_T utilizando la matriz de varianzas y covarianzas HAC (que puede ser la de Newey West, robusta a Heterocedasticidad y autocorrelación) θ_I : $\hat{V}(\hat{\theta}_I) = T \hat{\text{Var}}(f_T(\hat{\theta}_I))$. Por último, definimos $A_T = \hat{V}_T(\hat{\theta}_I)^{-1}$ como la nueva matriz de ponderaciones, y obtenemos θ_{II} . Este será asintóticamente eficiente. Intuición de por qué deberíamos usar lo anterior: Pensando solamente en la diagonal principal, y tenemos mucha precisión en algún momento (varianza muy pequeña), como estamos definiendo la inversa como matriz de ponderación, plantea que demos mayor peso a los instrumentos que son más precisos.

En este caso, no es necesario hacer más rondas pues las propiedades de la primera ronda son lo suficientemente buenas que hacen que no valga la pena seguir (de hecho, pueden empeorar si se hacen más).

Si se tienen instrumentos débiles, se infla la matriz de varianza y covarianza y por tanto, la estimación se hace imprecisa.

5.6 Inferencia

Note que la función objetivo multiplicada por T, tiende a una χ^2_{q-p} , donde q son las condiciones de momentos y p son los parámetros a estimar.

$$J_T = T Q_T(\hat{\theta}_T) \rightarrow \chi^2_{q-p}$$

Luego, si tenemos p parámetros a estimar, sería bueno tener $p + 1$ condiciones de momentos, para hacer que la distribución tenga la menor cantidad posible de grados de libertad. Además, siempre es mejor tener pocos instrumentos, pero relativamente buenos. Ahora bien, si se rechaza la nula quiere decir que se están rechazando las condiciones de ortogonalidad que se impusieron y por lo tanto, se debería rechazar la especificación del modelo, o sus supuestos.

Se podría hacer un test LRT, el cuál tendría la siguiente distribución

$$T(Q_T(\bar{\theta}) - Q_T(\hat{\theta})) \rightarrow \chi_r^2$$

donde r es el número de restricciones que se han impuesto.